

**UNCLASSIFIED**

---

**AD 297 381**

*Reproduced  
by the*

**ARMED SERVICES TECHNICAL INFORMATION AGENCY  
ARLINGTON HALL STATION  
ARLINGTON 12, VIRGINIA**



---

**UNCLASSIFIED**

NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

63-2-5

297381

297 381

297381  
ADDC - TDR - 63 - 11

ASTIA Document No.

31 December, 1962

CATALOGED BY ASTIA  
AD NO.

**AUTOMATIC LANGUAGE ANALYSIS**

**J. P. Thorne**

**Indiana University  
Bloomington, Indiana**

**Contract No. AF 30(602)-2185**

**Final Technical Report**

**Prepared  
for  
Rome Air Development Center  
Air Force Systems Command  
United States Air Force**

**Griffis Air Force Base  
New York**

**ASTIA  
RECEIVED  
MAR 6 1963  
TISIA**

**ASTIA NOTICE:**

Qualified requesters may obtain copies of this report from ASTIA Document Service Center, Arlington Hall Station, Arlington 12, Virginia. ASTIA Services for the Department of Defense contractors are available through the "Field of Interest Register" on a "need-to-know" certified by the cognizant military agency of their project or contract.

RADC - TDR - 63 - 11

ASTIA Document No.

31 December, 1962

**AUTOMATIC LANGUAGE ANALYSIS**

**J. P. Thorne**

**Indiana University  
Bloomington, Indiana**

**Contract No. AF 30(602)-2185**

**Final Technical Report**

**Prepared  
for  
Rome Air Development Center  
Air Force Systems Command  
United States Air Force**

**Griffis Air Force Base  
New York**

**ASTIA NOTICE:**

Qualified requesters may obtain copies of this report from ASTIA Document Service Center, Arlington Hall Station, Arlington 12, Virginia. ASTIA Services for the Department of Defense contractors are available through the "Field of Interest Register" on a "need-to-know" certified by the cognizant military agency of their project or contract.

## FOREWORD

Investigation of the application of computer techniques to the analysis of the syntactic and semantic structure of natural language was begun in February 1960 under Air Force Contract AF 30(602)-2185. The project began with an investigation of the structure of an artificial language suitable for storage and retrieval of information (see First Quarterly Report, May 1, 1960), but turned immediately to the development of routines for the analysis of English preparatory to any translation of natural language materials into an artificial language. Since it was deemed an essential feature of any artificial language that each sentence should consist of a single subject and a single predicate, work was started early on clause boundary marking routines, work which has continued until the present; many successive routines have been hand-tested, and several programmed (currently for the IBM 709). In order to test such routines a taped corpus was essential; for this purpose the book Planet Earth by Karl Stumpff (Ann Arbor, 1950) was completely key-punched and taped. For further work on scientific English, an additional corpus of 45 articles (five each for the fields of Mathematics, Physics, Astronomy, Agriculture, Geology, Chemistry, Engineering, Biology, and Medicine), totaling about 120,000 words, was selected by standard randomizing procedures under the direction of Professor Douglas Ellson of the Indiana University Psychology Department. A list of these articles is presented in Appendix XVII. Since the routines have not yet been adequately checked out on the preliminary corpus, this random corpus has not yet been punched.

Since any syntactical routines (including clause bracketing) required detailed form-class information on all words in the text, a complete dictionary of Planet Earth was prepared with form-class information coded for each entry.

As the routines have changed, the coding requirements have also been revised, and we are not prepared to state that our present coding is even yet all that is needed.

Since two of the requirements of an artificial language was originally judged to be minimum vocabulary (elimination of synonyms and derivatives) and automatic coding of as many words as possible, work was also begun early on a suffix-stripping routine which would automatically convert such words as "inclusion" into "include + N abstract" with all the codings automatically attachable to words formed with this suffix. (See Third, Fourth, and Fifth Quarterly Reports.) Work on this portion of the project has been temporarily shelved to deal with higher-priority projects; but it is hoped that eventually this can be taken up and prefixes (such as un-, re-, etc.) also dealt with in a similar manner. A much reduced and altered version of the affix routines has, however, been developed for use with the clause routines and the ambiguity routines. The need for these latter routines was also seen very early; since form-class assignments are required for clause-bracketing, and since many such assignments are ambiguous in English, routines were developed, and have been constantly improved, for eliminating such ambiguities.

To prepare the way for improvement of our suffix routines, the need was felt for a more complete reverse-alphabetized dictionary than was available in Walker's Rhyming Dictionary (the only published lexicon with this arrangement) and a program was written (and is available) for the reversing of Merriam-Webster's New Collegiate (1959), and five print-outs of this alphabetization were made (of which two have been forwarded to RADC). This work was done before the similar project carried on at the University of Pennsylvania, and it is believed that the two dictionaries in large measure complement each other.



Some hope was felt at the outset that part of the laborious hand-coding of dictionary entries might be avoided if a type of "machine learning" procedure could be adopted in place of algorithms, and so Douglas Elson conducted an experiment (see Appendix XVIII which showed (in our opinion, conclusively), that, while machine performance in syntactic analysis could be quickly improved up to a certain point, this part fell far below the requirements of information retrieval, at least with published dictionary coding only.

Work on semantic analysis has taken two slightly different directions at various periods, and only very tentative preliminary results are available as yet. The first approach was to consider the dictionary coding of certain semantic relationships (other than those conveyed by affix routines), such as synonymy, antonymy, hyponymy, etc. The most recent approach has confined its attention to devising mathematical measures of the degree of synonymy, using such available reference works as Roget's Thesaurus, Bartruff's Vocabulary Builder, and Morrison-Webster's Dictionary of Synonyms and Antonyms; future work is envisaged using such larger dictionaries as Morrison-Webster's New International (Third Edition).

Closely linked with the synonymy measure is the development of the syntactic agreement measure derived from the special arrangement and ranking of sentence words called FLEX, which is fully described in this report. Much of the work on FLEX and synonymy measure is due to Marion V. Cook and Donald Michie.

Work was done for a time (largely by J. P. Thorne, George Weolley and Jen Stewart) on the possibilities of artificial languages having specific logical properties (see the Sixth Quarterly Report). This was temporarily shelved when work of FLEX seemed to prove it a more interesting tool for information retrieval.

Most of those who have contributed in one way or another to this report have been named above; some portions have been written by Fred . Householder, Jr. Many parts of the report are indebted in various ways to work done by all the workers named in the various quarterly reports. The arduous task of preparing the appendices accompanying this report has been largely carried out by R. V. Cook.

## Abstract

Under the title Automatic Language Analysis (originally Automation of General Semantics) research has been directed towards the development of a device which would process and store completely unedited English language texts and print out answers to questions regarding these texts presented to it in their natural language form. The approach followed requires that the computer itself syntactically analyse input text in order to convert it into a special form called FLEX, which preserves only that syntactic information which is useful for data retrieval purposes. In their FLEX forms sentences can be compared to determine the degree of their relationship to each other in respect to both word-meaning and propositional meaning. A high correlation between a text sentence and a question indicates that the text sentence is a relevant answer.

The problems of construction a device of the kind proposed are discussed in some detail. The report also describes an approach to the mechanical analysis of language. It also contains an account of a version of FLEX and of certain preliminary experiments in weighting syntactic information. The fourth section briefly lists the conclusions and recommendations, both positive and negative, which may be drawn from the work of the project.

## TABLE OF CONTENTS

I	General Discussion	1
II	1. Construction of a Device for Syntactic Pattern Recognition in English	11
	2. Construction of a Routine to Resolve Ambiguous Form-Class Assignments	18
	3. Computer Technique for the Solution of Form-Class Ambiguities	21
	4. Construction of a Routine to Make Clause Division in a Partially Analyzed Sentence	40
III	1. FLEX, A Propositional Form for Information Retrieval	48
	2. Outline of Information Retrieval Process	64
IV	Conclusions and Recommendations	66
	BIBLIOGRAPHY	69
	APPENDICES	70
I	Card Formats	70
II	File Formats	71
III	Block Diagrams - File Development	73
IV	Analyzable Unit Schematic	74
V	Phrasification Flow	75
VI	Form Subclass Codes	76
VII	The Dictionary Codes	79
VIII	Affix Dictionary Flow	80
IX	Control Flow	82

# Table of Contents (Cont'd)

X	Indicator Flow	83
XI	The Indicator Codes	84
XII	The Search Routines	90
XIII	Typical Ambiguity Rule	92
XIV	The Ambiguity Routines	93
XV	Absolute Breaker Lists	123
XVI	Schematic Representation of Clause Division Rules	127
XVII	A Randomly Selected Scientific Corpus	134
XVIII	An Intuitive Computer Learns an Element of Grammar	152
XIX	Concerning the Matching Formulae	167

## General Discussion

The work reported upon here is intended as a contribution to the study of the problem of direct communication between men and computers through the medium of natural language. It deals with the design of a device which would efficiently store a large body of data fed into it in the form of completely unedited natural language texts, and answer any question posed to it in natural language regarding these texts. The construction of programs to carry out certain procedures which would form a necessary part of the operation of such a device are described in some detail. It could be agreed that, as they stand, these procedures - indeed, the choice of certain of these procedures for development rather than others - represent fairly drastic compromises when compared with what might be regarded as the ultimate goal in this field of research. This is true. The model described here is intended to satisfy the minimum requirements that a priori can be demanded of such a device if it is not to produce insignificant results. Its purpose is mainly heuristic. Lacking such a device it seems difficult, if not impossible, to see how better results could be obtained. Its realization would establish a context in which empirical experiments could be conducted. Examination of the results of these experiments would suggest the best ways in which improvements could be made.

The approach followed differs in certain important respects from other information retrieval experiments. While obviously having much in common with mechanical document retrieval systems it differs from them in aiming at entirely dispensing with the human abstracter who prepares a specially organized text which is then translated into a machine language for transmission to the computer; documents are retrieved through the recognition of certain key words or of a combination of key words occurring in the stored abstract and the information request.

Although sharing many of the problems encountered in developing techniques for the production of mechanical abstracts it differs from this work particularly in the matter of scope. The intention here is to enable the machine not to select from a text those parts which, in accordance with certain built-in criteria, are rated as most significant, but to handle the whole text. This would remove the risk of the total loss of some information due to the machine having to act under the limitation of restrictive pre-judgments concerning what is to be considered important in certain kinds of texts - something which must obviously change with different people and different times. It also means that a very much higher standard of precision can eventually be expected. The computer will not merely supply a list of references to documents which contain data relevant to a particular request for information. If problems of storage space can be overcome the machine should be able to print out the actual data itself. Failing this it should be able to provide something as precise as page and line references. Without ceasing to be useful for information retrieval purposes the computer can become a 'question-answerer'.

An automatic question-answerer, Baseball, is already in operation at M. I. T.<sup>1</sup> The computer stores information from a restricted field (the date, place and score of every game played in the American League in 1959) fed into it in the form of a special machine language. Questions regarding this data, phrased in natural language (English) and ranging over a fairly wide degree of lexical and syntactical variety, are automatically answered. The outstanding feature of the device is its capacity to carry out quite complex procedures which enable it to answer questions of a considerable degree of complexity.

A question like, "Is it true that every team played at least once in each ball-park during the month of August?" for example. In this particular

---

1. See 'Baseball: An Automatic Question-Answerer'. Bert F. Green, Jr., Alice K. Wolf, Carol Chomsky, and Kenneth Laughery Proc. WJCC, Los Angeles, Calif. May 9-11, 1961

respect Baseball goes beyond the kind of device proposed here. It differs from it also in two other respects. First, although the machine automatically processes the questions put to it, the data it stores is fed into it in an already processed form. Second, the data it operates upon is of a very special kind. The model for the device appears to be a quiz-competitor. The machine answers rightly or with "don't know". The device we propose is modelled upon more usual human behavior. Under ordinary circumstances the giving of a direct answer to a question is fairly unusual. More typical is the phenomenon psychologists call "answer by withdrawal". That is to say, when questioned one tends not to answer the questioner directly but by presenting him with the information at one's disposal which seems most germane to his query. It is this facility that we wish the computer to simulate. The advantages of building this capacity into an information-retrieval device and the difficulties of accomplishing this aim are now discussed.

Consider the case of a computer with the following sentences stored in its memory: (i) The boy hits the ball and (ii) The girl drinks the coffee. It is desired to program the computer in such a way that in response to the question Does the boy hit the ball? Sentence (i) and not Sentence (ii) will be printed out. It is assumed that no system of labeling sentences is to be employed. The machine is to interpret the question itself as an instruction to search the memory for the sentence which matches it most closely. The question provides a pattern the elements of which are the individual words. A search is initiated for patterns which coincide with that contained in the question. This is the simplest information retrieval situation involving a natural language text and a natural language question. Notice that already we have to allow for a certain degree of tolerance in matching patterns. If



only exact matches are to be accepted then Sentence(i) will not be retrieved since it contains one less word (also a different punctuation mark if these are to be counted). This must obviously be allowed for. We would wish sentences like (iii) The boy hits the ball hard and (iv) The boy hits the ball every time to be retrieved in response to this question if they formed part of the corpus. Notice also that the sentences (v) The boy never hits the ball and (vi) The boy did not hit the ball would also be considered significant answers. In matching patterns the occurrence of negative elements can be ignored. Negative answers are as likely to be relevant as positive answers. A failure to retrieve any answer at all is to be interpreted as "don't know", not "no".

Now consider a slightly more complicated situation. We substitute for Sentence (i) Sentence (vii) The boy strikes the ball. The same question is posed. If we specify that it is the sentence which offers the closest match that is to be chosen Sentence (vii) will be printed out and not Sentence (ii). But now the notion of tolerance has to be considerably extended. The match is made on the recognition of the common pattern The boy---the ball. This would mean that the sentence (viii) The boy kicks the ball (ix) The boy avoids the ball and (x) The boy recognizes the ball if added to the corpus would be retrieved at exactly the same time as Sentence (vii). More disturbing still, if (xi) The boy enjoys the ball is added to the corpus it too will be retrieved as a relevant answer.

It seems then that The boy---the ball cannot be regarded as a sufficiently well-defined pattern for its occurrence in the question sentence and the corpus sentence to be taken as a criterion for the retrieval. This decision prevents the noise sentences being printed out. It still leaves the problem of retrieving Sentence (iii) which on all counts is a relevant answer to the question.

The solution adopted demands that strike be considered as being an element of the same type as hit. There is, of course, strong intuitive motivation for this. The notion is reflected, though perhaps not very precisely, by an entry in a thesaurus - a list of words all substitutable for at least one other word in the list in some context or other. We have postponed discussion of the important question of the construction of a thesaurus for purposes of information retrieval from natural language. For the moment we assume that the computer is also supplied with a dictionary and that against each entry is a list of numbers corresponding to those heading thesaurus entries in some standard work like Roget's Thesaurus. A look-up routine now reveals that hit and strike have an identical entry. The question sentence and Sentence (vii) are now treated as sharing the same three-element pattern. Sentence (vii) is chosen and the rest rejected.

Identical listings in a thesaurus are rare and in an ideal thesaurus probably non-existent. On the other hand partial overlappings are frequent. It is reasonable to suppose that the words hit and kick, for example, would occur together in at least one thesaurus cluster. An extension of the present method would interpret the <sup>detection</sup> of this overlap as indication that kick also fills out the pattern but in a less exact manner. So that The boy kicks the ball is printed out in response to the question Does the boy hit the ball? but only after The boy strikes the ball has already been retrieved. These machine responses can be equated to the replies "Yes" and "No, but he kicked it". The point is that the second of these might be relevant information. Whether it actually is or not can only be decided by the operator, not the machine.

A further improvement is suggested by the observation that in Roget's Thesaurus the principle of arrangement is such that a list of terms comprising

one entry is likely to be immediately followed by one made up of a list of their opposites. The fact that a negative answer is likely to be as relevant as a positive one suggests that in some cases adjacent clusters should be merged. This would bring hit and miss together in the same cluster, and since it can be assumed that miss and avoid also occur in the same list, this extension of the notion of negative responses means that Sentence (ix) will also be printed out.

Let us assume that each time the machine recognizes a match between the lexical items in a question and those in a stored sentence it makes a simple computation. Let us say that it scores 1 for a full match hit: Hit or Hit: strike and  $\frac{1}{2}$  for a partial match hit: kick, hit, avoid. The search for matches now produces a rough kind of ordering of the sentences in the corpus in terms of their relevance to the question. Ignoring words such as definite articles which occur in no thesaurus entries Sentences (i) and (vii) score 3, Sentences (viii) and (ix)  $2\frac{1}{2}$ , Sentences (x) and (xi) 2 and Sentence (ii) 0. If the program were designed in such a way as to ensure that the sentences were printed out in descending order there would be no need to decide beforehand on a cut-off point, i.e. a score below which no sentence is printed out. Once the operator observes one sentence in the print-out that he regards as irrelevant he can stop the production of output on the assumption that the rest of the information contained in the corpus is irrelevant to his particular need.

Increasing the number of potential matches by associating with each word a list of thesaurus entries does not remove all the problems. In fact it introduces new problems. Consider the Sentences (xii) The well-behaved girl sips her coffee and (xiii) Coffee is good to drink and the question Does the young lady drink coffee? Computing matches in the way previously suggested

we find that both text sentences make the same score (xii) (young) lady: girl= $\frac{1}{2}$ , sips: drink= $\frac{1}{2}$  coffee: coffee 1 Total 2, (xiii) coffee: coffee= $\frac{1}{2}$  drink: drink=1 Total 2. Thus although the system enables (xii) which has only one exact match to score the same as the irrelevant(xiii), which has two, it fails to establish any priority for its retrieval. It also means that the noise sentence (xiv) A taste for liquid honey is characteristic of immature female bears, where taste, liquid, immature and female all score  $\frac{1}{2}$  points, would also be retrieved. The difficulty cannot be overcome by making a rule that matching patterns must not only contain the same words but the same words in the same order. One would obviously want to retrieve (xv) Coffee is drunk by all well-bred young ladies, which would be rejected by such a rule.

To overcome this difficulty we advance the hypothesis that the lack of discrimination in the information-retrieval systems so far developed arises from the fact that they utilize only semantic information and fail completely to make use of syntactic information. They fail to recognize that (to put it the simplest fashion possible) Sentence (xiii) is "about" coffee and its relationship to drinking. It is reasonable to claim that this is reflected in the statement that (to use the most neutral grammatical terminology available) the subject of (xii) is girl, while the subject of (xv) is coffee. Sentences which have words, or words from the same thesaurus cluster, in common are more closely related from our point of view than those which do not. But the factor which controls how closely these sentences are related is the extent to which the common elements share the same places in their syntactical pattern. Matching procedures based on syntactic information can be made extremely comprehensive. A match between the subject of Sentence (a) and the subject of Sentence (b) is obviously more significant than a match between the subject of Sentence (a) and

the object of Sentence (b), but this second match is still important. It is more important than say a match between the subject of Sentence (a) and a modifier of the object of Sentence (b). Yet this fact too might not be without significance and provision for it can be made in the system.

The question of the explanatory power of grammars has only recently begun to interest linguists. Its importance is a direct outcome of the decision taken by certain linguists to make the goal of syntactical studies not the elaboration of procedures by means of which linguistic data can be classified, but the construction of grammars which are their theories of sentence structure.<sup>2</sup> Since the observable data such a theory has to cover is infinite it must be generative. A grammar of this kind is structured as a calculus which generates all and only the well-formed sentences of a language. Given a complete grammar of a language, for any well-formed sentence of that language there should exist in it an ordered set of formulae which would generate this sentence. This is expressible as an ordered (bracketed) string of syntactic symbols and constitutes a syntactical analysis of the sentence. Since there is no limit on the number of generative grammars that can be constructed for a language the question of choosing between competing grammars becomes crucial. It has been suggested that the most powerful grammar is that which produces analyses which most fully explain the native speaker's intuitions about the relationships (sameness and differences) between sentences. Preliminary investigations into the account of syntactic information

---

2. This trend is particularly associated with the work of Noam Chomsky. A clear statement of his views is given in 'The Logical Structure of Linguistic Theory', Reprints of Papers for the Ninth International Congress of Linguists (Cambridge, Mass. 1962). See also M. A. K. Halliday "Categories of the Theory of Grammar", WORD, Vol. 17 (1961).

necessary for efficient data retrieval from natural language texts suggest that, since only relatively coarse distinctions between the syntactic functions of matching words appear to be significant, a comparatively weak grammar will prove adequate.

The adequacy of weak grammatical models for information-retrieval purposes is important for two reasons.

(i) The complete grammar which linguists regard as the ultimate goal is unlikely to come into existence for a considerable time. At the moment we do not possess a single grammar for a natural language which comes anywhere near to satisfying demands of completeness that might reasonably be made of it when no other purpose underlies its construction than the study of language itself.

(ii) The linguist who constructs a generative grammar is interested in data only in so far as it tests his theory. Given a sentence which a native speaker agrees is well-formed he is concerned only that his grammar should contain an ordered set of formulae that will generate that sentence. The question of how he decides which set of formulae in the grammar satisfies this condition is of no interest to him. It is not part of the theory. A grammar does not tell one how to recognize the analysis of sentences. The process of checking an "ideal" sentence generated by the grammar and the actual sentence observed affords no difficulty to human beings. Human beings have intuitions about language anyway and an illuminating grammar is illuminating precisely because it in some sense incorporates these intuitions. We want to enable the machine to make use of grammatical information because it does not have intuitions about language. The problem of how a machine is to determine the syntactic analysis of any sentence, is, therefore, crucial. The application of

computers to the task of data-retrieval from natural language demands the construction of a heuristic device which will enable it to derive from any sentence an analysis which (a) corresponds to an ordered set of formulae in a grammar (b) meets the strong intuitive demands of fitness when applied to the sentence from which it has been derived. A similar device is essential to machine translation. Machine translation demands reference to an extremely powerful grammar. A characteristic of the weak models useful for information retrieval purposes is the relatively small number of formulae they contain. The problems of construction of a recognition device for this grammar are thereby greatly eased.

Any really significant reduction in the complexity of syntactic analyses, however, is not as easy to obtain as it might seem. It is an inescapable fact that if a recognition device is to work at all efficiently it will almost certainly produce analyses containing more information than can be used for data retrieval purposes. The reason for this is simple. Even if all that is demanded of a recognition device is that it should indicate nothing more than the subject, verb and object of a sentence, it can only do so as the result of first assigning labels to all, or almost all, the elements in the sentence. The problem here is not just that certain words can be either nouns or verbs, or verbs or adjectives, etc., but that any noun can be either a subject or an object or the modifier of a subject or object, or the modifier of a modifier, etc. The last difficulty arises out of the recursive tendencies of natural language. Sequences of the kind, the boy with the dog with long ears with white spots, etc. are possible. Generative formulae for this might be Subject  $\rightarrow$  Nominal Group  $\rightarrow$  Nominal Group + Nominal Group  $\rightarrow$  Nominal Group + Nominal Group + Nominal Group. We may be concerned only with the identification of the element

that enters into the first of these formulae, but this is impossible until the structure of the whole string has been established.

To avoid the danger of being weighed down by redundant information we propose that the initial processing both of data and questions consist of a translation into an artificial language in which only those syntactic relationships which prove significant in information retrieval occur. The analyses produced by the recognition device are not stored but interpreted as a set of instructions for rewriting the sentence into the equivalent form in the artificial language. In this way we are relieved of the problem of constructing a complete grammar for English. It is only necessary to ensure that every analysis will contain sufficient information to enable the sentence for which it has been produced to be rewritten as a well-formed sentence in the artificial language. To ensure the completeness of the grammar of the artificial language is a relatively simple matter. The artificial language will contain only a small number of sentence types to which the large number of English sentence types will have to correspond. Artificial languages for information-retrieval differ, therefore, in a very important respect from those created by logicians. Instead of reducing ambiguity they promote it. Certain forms (actives and passives for example: The boy hit the ball, The ball was hit by the boy) are related syntactically: a fact that can be incorporated into a grammar, but only at the cost of some added complexity. In the "collapsed" grammar of English we propose, it is merely necessary to ensure that both forms are rewritten identically in the artificial language.

Marching procedures are carried out on sentences in their translated forms. For ease of reference the artificial language has been given the name FLEX. Since its vocabulary consists of thesaurus clusters rather than words it is characterized by minimum syntactic organisation and maximum semantic organization. It differs most noticeably from natural languages in not dis-



playing any of the syntactic variations which prevent the development of what Yngve calls "depth."<sup>3</sup> In its present form nearly the whole grammar of FLEX is contained in the formulae.

- (1) S → Subject + Predicate
- (2) Subject → Noun
- (3) Noun → (modifier) + Noun
- (4) Predicate → Verb + Object
- (5) Verb → (Modifier) + Verb
- (6) Object → Noun

The formulae (3) and (5) are recursive, it being possible to add an infinite number of modifiers to subjects, verbs and objects though in practice a limit is set at four. The rules are not ordered; (3) is applied again after (6).

This particular language is almost certainly too simple to promote efficient data retrieval. Its usefulness at the moment consists in producing a context in which hypotheses concerning the adequacy of various types of thesauri and the weighting of syntactic information can be readily tested. Some experiments in texting various functions for weighting syntactic matches are described in Part III.

From another point of view FLEX may be regarded not as a language, but as a device for (a) splitting each (simple) sentence of a natural language into two parts, "Subject" and "Predicate", and (b) assigning weights to each word in each part according to their "importance". The assumption is made that, in general, words which are grammatically superordinate, or "head" words, will be more important for information, and modifiers will be less so, while such items as conjunctions, particles, prepositions, articles, and the like are of no importance whatever.

## II. 1. Construction of a Heuristic Device for Syntactic Pattern Recognition in English.

The need for a device which will assign a significant syntactical analysis to any English sentence was explained in part I. A syntactical analysis is defined as an ordered sequence of symbols, all of which occur in the same grammar, and in which each symbol corresponds to one word in the

---

3. V. H. Yngve 'A Model and an Hypothesis for Language Structure, Proceedings of the American Philosophical Society, Vol. 104 (1960), pp. 130-8.

one word. A grammar is defined as a calculus of formulae which generates all and only all the well formed sentences of a language. A correct syntactical analysis is defined as an ordered string of symbols which would be generated by a grammar. A significant analysis of a sentence is one which is correct and which meets strong intuitive demands of fitness when assigned to that sentence.

An idea of the problems encountered in developing such a program can be gained by considering the simple three word sentence These points stand. The sentence is syntactically unambiguous (Demonstrative A<sub>d</sub>jective + Noun) + Verb, being the only significant analysis. A computer analysis, however, is the output of look-up routines operating with a dictionary which lists for each word in it all the syntactic roles that can be undertaken by that word. (An account of the structure of this dictionary and the look-up routine will be found in Section 2 of this report and associated appendices.) Since it is uneconomical to enter both boy and boys, jump and jumped, jump and jumping etc., in the dictionary, and since also these suffixes contain essential syntactic information, it is necessary for the look-up routine to work in conjunction with a routine which recognizes boys as boy + s and interprets the s as indicating plurality in the noun. This routine is also described in Section 2. In the present case the look-up routine would find two entries for each word. These -- Demonstrative Adjective and Demonstrative Pronoun, Points - Noun and Verb, Stand, - Noun and Verb. Applying a rule which states that each word can fulfil only one syntactical function at a time in an analysis results in a total output of eight different unranked strings.

The demand that the machine should select only the significant analyses imposes too strong a condition. It is sufficient if it retains only those which are correct. In practice it can be assumed that any output that can be shown to be a correct analysis will also be a significant analysis of the sentence from which it is derived. A maximum program, therefore, would guarantee the correctness of all analyses printed out. The difficulties of attaining such an objective are perhaps insurmountable. In its place we offer an approach which has as its objective the recognition of nearly all incorrect analyses. In this connection we offer the tentative hypothesis that we select one syntactical analysis of a sentence over all other possible analyses not so much because it is the right one as because it is the only one that is not wrong. In the case where there are two right analyses the sentence is syntactically ambiguous. In the same way if the machine could recognize incorrect analyses then it could be assumed that any not rejected on this count would be correct.

It is impossible to supply the machine with a list of incorrect analyses. There is no reason to suppose that such a list would be any shorter than a list of correct analyses. Probably both are infinite. The solution we adopt arises out of three observations:-

(1) The most frequent and intractable cases of ambiguity of form class assignment seem (at least in English) to occur when one of the assignments is verb.

(2) A study of analysed sentences indicates that an operational definition of a clause as that part of a sentence which contains one and only one verb (excluding auxiliaries and modals when these are followed by other verbs) is quite workable.

(3) In the course of constructing a generative grammar many of the greatest problems are encountered in the area covering the generation of sentences comprising more than one clause. It is here that problems of selection - that is problems of deciding how far the individual strings making up each clause should be developed before they are associated in the course of generating the complete sentence - appear to be particularly troublesome. The difficulty of finding the simplest theoretical statement indicates that there are a large number of specific features to be taken into consideration. This in turn suggests that a list of incorrectly juxtaposed clauses, described in terms of these features would provide criteria for rejecting nearly all incorrect analyses. This would also cover one-clause sentences since the number of clauses which can combine with the null class is strictly limited.

The procedure we propose for mechanical recognition of syntactical analysis consists of six parts.

1. Formation of a list of syntactic symbols for each sentence by dictionary look-up.

2. Resolution of ambiguous assignments where one of the assignments is verb. Accepting the operational definition of clauses given above, this will mean that at the conclusion of this operation the analysis will contain sufficient information for

3. Ordering into clauses to be imposed on the string.

4. Checking against a list of incorrect clause combinations. Detection of an error here must, by definition, mean that a mistake has been made earlier in the resolution of an ambiguous assignment involving the form class verb. In this case we enter

5. The rules employed for resolving form class ambiguities in 2

are arranged in descending order according to their efficiency. The number of the rule used to resolve an ambiguous assignment is carried forward with its output. In this way once a mistake has been detected at the clause level candidates for re-assignment are easily recognized. In cases where there is more than one candidate the one with the highest rule number is changed first. The process is continued until a legitimate result is obtained.

This simplified scheme may be improved through re-entry into the resolution cycle. At this level we may discover that a word group taken as an idiom by the phrasification does not in a particular instance operate as a unit. Such a phase would have to be dismantled before re-entry. Also prepositional phrases which act as adverbs can be so noted; and all words resolved by "perfect" rules can be noted as unambiguous. Such information would obviously strengthen the verb resolution of phase 2.

6. Ordering into phrases within the clause and the resolution of the remaining ambiguous assignments are carried out.

The number of rules need in 2 is kept down to a comparatively small number by adopting the strategy of giving all uncertain cases the assignment "verb." This means that in the instances where this is wrong an extra clause is interpolated into the sentence, making the chances that a correct analysis will be formed very slight and ensuring a later correct re-assignment.

Programs for Parts 1 and 2 have been written and tested. Work on part 3 is nearly complete. Work on part 4 has been postponed until all the earlier ones are fully operational. The reasons for this are partly that the results obtained from the earlier routines are so good that it is

possible that certain incorrect juxtapositions of clauses can be ignored as impossible outputs of the earlier routines. Investigations so far made indicate that the hypothesis, that, if a description (even though partial) of the clause level stages in the generation of a sentence can be derived from it, then this will contain sufficient information to cause most incorrect analyses produced by the look-up routine immediately to be rejected, is a valid one. So far only a handful of cases have been discovered where wrong analyses would be derived which would not produce clause juxtapositions that one would expect to find in the list of incorrect combinations proposed above. An example of how this will work is provided by the sentence:

'One of the most satisfactory laboratory experiments in the field of mechanics is the measurement of surface tension by means of a Du Nouy tensiometer.'

The look-up routine would produce the following string

Adjective/Pronoun + Preposition + Definite Article + Adjective/Adverb  
 + Adjective + Noun + Noun Plural/Present Tense Verb + Preposition  
 + Definite Article + Noun/Present Tense Verb + Preposition + Noun  
 + Auxiliary Verb + Definite Article + Noun + Preposition + Noun/Present  
 Tense Verb + Noun + Preposition + Noun/Present Tense Verb + Preposition  
 + Indefinite Article + Noun + Noun.

The ambiguity routines would resolve all the ambiguous assignments correctly except in the case of the word experiments which is wrongly assigned as a verb. The resultant string is now

(Adjective/Pronoun + Preposition + Definite Article + Adjective/Adverb  
 + Adjective + Noun + Verb + Preposition + Definite Article + Noun

+ Preposition + Noun) (+ Auxiliary Verb + Definite Article + Noun  
 + Preposition + Noun + Noun + Preposition + Noun + Preposition  
 + Indefinite Article + Noun + Noun.)

The routines for dividing the sentence itself into clauses would break up the string into two clauses in the manner indicated. Routine 4 would indicate that the juxtaposition of a clause containing a present tense verb and no marker of subordination (which, who, etc.) and a clause headed by an auxiliary verb is incorrect. Routine 5 would reveal experiments, field, surface and means as possible candidates for re-assignments. Experiments is the one resolved by the highest numbered (i.e. weakest) rule. Its assignment is changed to noun. Clause analysis on the new string now produces only one clause. Since this clause can stand by itself to form a whole sentence, no listing for clause containing present tense verb and no marker of subordination with the null class is found by Routine 4 and the new analysis is allowed to stand.

## 2

### Construction of a Routine to Resolve Ambiguous Form Class Assignments.

For reasons already explained it was decided in the first instance to construct routines to resolve only those ambiguities involving the possible assignment verb. This meant the construction of five separate sets of rules for the resolution of the ambiguities; noun/present tense verb (point, stage, face etc.), adjective/present tense verb (clean, complete, close etc.) noun/past participle (and sometimes present tense also) (cut, set, felt, thought), adjective/past participle (fixed, interested, given, etc.) and present participle/adjective/noun (meaning, using, running, etc.)

(See Appendix XIV, the Ambiguity Routines). It was subsequently decided to add a sixth set of rules in which the ambiguities involved in words of such idiosyncratic distribution as like, except, might, can, will, even, still, well, and a few others would be resolved.

A rule is an instruction to search the environment of an ambiguous item for the presence of another item or items which are diagnostic in this context. For example, in each set of rules one of the first to apply is that which initiates a search for a definite article immediately in front of the ambiguous item. If this is found the ambiguous symbol in the output of the look-up routine is rewritten as noun. If the particular diagnostic item cited in the rule is not discovered the next rule is applied and the search for another diagnostic item started. Since the input for these routines consists of the form class assignments for each word in the sentence read from left to right, the machine resolving each ambiguity as it comes to it, most rules direct searches to the left hand side of the environment; the right hand side being likely to contain only information which is itself ambiguous. In the interests of a simple machine solution very few rules demand a search extending over more than three items in either direction. Since the distribution of adverbs in English is extremely wide they rarely serve any diagnostic function and most rules include the instruction not to count adverbs as part of the environment of an ambiguous item.

The ordering of the rules is important for two reasons. First, certain items are diagnostic only in the absence of some other diagnostic items. The rules must, therefore, be arranged in such a way as to ensure that a search



for the first set has already been undertaken and failed before the others are looked for in the environment. Second, since the general routine demands that a decision be made in each case, and since it is known that the rules will sometimes produce wrong results it is important that information as to how likely it is that the criteria used as a basis of a particular decision will produce mistakes should be readily available. In particular this provision enables us to exploit the strategy of resolving most uncertain cases as verbs. This in turn means that the number of rules in each routine can be kept down to a bare minimum.

Many items are used in more than one set of rules. The definite article, for example, is diagnostic when it occurs in the environment of an item designated by the look-up routine as noun/present tense verb, or an item designated adjective/present tense verb, or as adjective/past participle, as noun/past participle, or as present participle/adjective/noun. Many items regularly fulfill the same diagnostic functions as others. The discovery of an indefinite article in the environment of an ambiguous item nearly always leads to exactly the same result as the discovery of a definite article. A great economy is attained in the set of the machine solutions by assigning to each word an indicator code which indicates all the diagnostic functions it can fulfill. As each word enters the routine it is considered from two points of view; first to see if it is ambiguous and, if so, which kind of ambiguity it displays, second to see if it can play any part in the resolution of the ambiguous form ~~class~~ assignments of other words, i.e. which indicator codes belong to it. A description of the programming of these routines follows. The full set of the rules in the forms most convenient for programming and hand checking will be found in Appendix XIV to this section.

## Computer Technique for the Solution of Form Class Ambiguities

### Data Preparation

The raw material used in the computer solution to the word ambiguities problem can be divided into two categories: dictionary data, and scientific text data.

### Dictionary Data

The former category was originally prepared as a tape file for use with the IBM 650 computer programs. This dictionary tape file has been converted by an IBM 7090 routine<sup>4</sup> to a format which is suitable for the IBM 1401 and 709. At the time of its conversion, the dictionary file was modified both in content and in structure to its present format. (See Appendix II, File 1.) The preparation of this category of data is, therefore, complete. Additions, deletions or changes to the file will hereafter be made through the updating program which will be described in a later section.

### Text Data

The initial scientific text, Planet Earth<sup>5</sup> was also prepared as a tape file for use with the IBM 650 Computer programs and it also has been converted by the above-mentioned routine<sup>4</sup>. See Appendix II, File 2. A special IBM 709 program is being written which will be used for assigning to words of this text only the text identification described below.

The preparation of scientific text data is a continuous part of this data processing system developed for this project. A description of the manner in which this data is prepared follows.

Text data preparation is carried out in two stages: the conversion of the printed material to punched cards; and the conversion of the punched card to tape files.

4. D. E. Flanigan, IBM. Tape Conversion Routine for the IBM 7090.

5. Planet Earth. Karl Stumpff (Ann Arbor, 1959).

### Text-to-card

The text is punched (See Appendix I, Format 2) in much the same way as it is typed. The following conventions are observed in the key-punch operation.

1. Words and punctuation are separated by blanks.
2. Every paragraph and every page begins a new card.
3. When a new word begins a card, the first column is blank.
4. When a new paragraph begins a card, the first two columns are blank.
5. When a new page begins a card, the first three columns are blank.
6. When both (3) and (5) occur, the first four columns are left blank.
7. When both (4) and (5) occur, the first five columns are left blank.
8. A sentence or a paragraph may be continuous over many cards.
9. The last 8 columns are reserved for card sequencing.

### Card-to-tape

The primary purpose of the card-to-tape conversion operation is to provide an efficient form of input to the data processing system. Its function is to provide each text item with an identification number corresponding to its place in the text.

The conversion is to be done on the IBM 1401. The logical records produced contain the text time, with the capital marker moved to a convenient location, the identification of the text and the identification for each item in the text. Punctuation is moved over one character in its record.

The format of the 10-character text identification number is:

PPP P SS C U II

where

- PPP is the page number in the text.
- P is the number of paragraph beginning on the page.
- SS is the sentence number within the paragraph.
- C is the clause number within the sentence.
- U is the analysable unit within the sentence.
- II is the item number in the sentence.

The convention is followed that a paragraph is considered as being all on the page on which it begins. If a paragraph is more than a page long the beginning of the next paragraph has a page number which is greater by (at least) two than the page number of the preceding paragraph.

Unit numbers are assigned later by the unit analysis routine. Clause numbers will be assigned by the clause analysis routines.

#### File Standards

For purposes of uniformity within the data processing system, all files conform to the following standards:

1. Tapes are written at low density;
2. Logical record length within any file is fixed;
3. Block (physical tape record) length within any file is fixed;
4. Data files are all BCD files;
5. In addition to its data blocks, each data file contains one header record and one trailer record as described in the IDCS manual.<sup>6</sup>

---

6. IBM Reference Manual C28-6100-1, 709/7090 INPUT/OUTPUT CONTROL SYSTEM, pp. 22-23.

(6) Each blank tape mounted for use as output of the 709 computer programs contains the Blank Tape Label described in the IOCS Manual.<sup>7</sup>

### File Maintenance

The data processing system for resolving word ambiguities required two types of maintenance programs: sort programs and update programs.

Sorting is carried out by means of the standard IBM 709 Sort.<sup>8</sup> Sort sequences are designated both in the file descriptions (Appendix II) and in the block diagrams (Appendix III).

Updating of all files of the system may be carried out by means of a single IBM 709 program which is currently being written.<sup>9</sup>

### Update Program

The particular function which the program performs at any one time depends upon the specifications stated by the user. These specifications include a general description of the characteristics of the file to be modified and of the manner of modification (Appendix I, Formats 4).

Specifications are punched on cards and converted to tape by a standard IBM 1401 program in which each card becomes one fourteen-word record. (See Appendix II, File 3). This "change" tape and the tape file to be modified are the two inputs to the file update program. The primary output of the program is the modified file. A secondary output file can be produced if requested in the specifications.

- 
7. IBM Reference Manual C28-6100-1, 709/7090 INPUT/OUTPUT CONTROL SYSTEM, p. 21.
  8. IBM Reference Manual C28-6036, 1959, Generalized Sorting Program for the IBM 709 DATA Processing System: Sort 909.
  9. Instruction files are being updated by standard Comptrand package update.

## Data Files

Data files are updated by means of cards #3 and #4. Depending on the use of the parameters in these cards the program may be directed to:

1) Locate a logical record by its position in the file, and add N logical records immediately after the one located.

The actual records added follow immediately behind the specification card #4. Each of the N records must begin in column 1 of a card; the record may be continued from card to card (78 columns per card) until the full size of the record has been punched. Continuation cards are identified by a "c" in column 79. Final  $\phi$  padding of records need not be punched. Since the record and block size are both stated in card #3, a full record will always be added.

2) Locate a record by its position in the file, and delete--beginning with that record or the following one--N records. No records need follow this specification card #4.

It is possible in updating data files to delete certain records and to add others at the same point in the file. This must be done by means of two specification cards #4, each referring to the record after which records are to be added. The specification card with the "Add" parameter and the records to be added must precede the specification card with the "Delete" parameter.

3) Locate a record by its position, and change that record of the file with the match field stated on the specification card #4. When a record is found in which the two fields agree, delete or add N records, or change that record by the replacement field of the specification #4.

4) Locate<sup>10</sup> a specified field in every record of the file. Whenever the field specified matches the field stated on the specification card #4, change the record by the replacement field of the specification card #4.<sup>11</sup>

The dictionary file (See Appendix II, File 1) may be updated by record in any of the above-mentioned ways. However, in addition, the logical records within the dictionary file which contain the count of the number of other records within an alphabetic grouping are modified in accordance with the addition or deletion of records during updating .

Data files may also be examined by means of the specification cards #3 and #4. When used for this purpose, the specification cards direct the program to locate a specified field in every record of the file. Whenever the record field matches the field specified in card #4, the record is placed on the secondary output file.

All specification cards #4 except those using the "Add" parameter may request secondary output. When requested in connection with deletions, the secondary output file will contain all records which have been deleted from the modified file.

When requested in connection with changes, the secondary output file will contain all records which have been modified on the primary output file.

- 
10. Neither the matching field nor the replacing field may exceed the size of the record. The match field must always be completely stated on the specification card but may continue from card to card until it is completely stated. No characters - including zeros - may be omitted from the field statement.
  11. By this direction all specification cards #4 are applied to each record of the file. This makes possible a change which is universal to all records of the file, or, if no change is requested, the selection of all records which have some common characteristic.

### Data Modification

There are two simple types of text which prepare it for further language analysis. The first allows the removal of obvious parenthetical expressions from the middle of a sentence for separate analysis. The second allows us to consider simple groups of words which normally function as one to be treated as a single item.

### Separation into Analyzable Units

This routine (See Appendix IV) is to set off pairs of parentheses and dashes and the items inclosed by them so that the item immediately following such a unit may be considered as immediately following the item which precedes this unit. This is done by assigning unit (analyzable unit) numbers to each item.

Such units can then be placed at the end of the sentence by a standard I.B. sort routine. Sorting on digits 1-6 and 8-10 sequences the file into "analyzable unit" order. Sorting on digits 1-6 and 9-10 would sequence the file into original text order. For example:

"It is necessary (for purposes of this analysis) to consider these items together." will be analysed:

It	Is	Necessary	(
ID0011010001	0011010002	0011010003	0011010104
for	purposes	of	this
0011010105	0011010106	0011010107	0011010108
analysis	)	to	consider
0011010109	0011010109	0011010010	0011010010
these	items	together	.
0011010011	0011010012	0011010013	0011010014



and this can be sorted, as above, into:

"It is necessary to consider these items together. (for purposes of this analysis)"

### Phrasification

To allow such phrases as "of course" to be treated as preposition + noun is to complicate the grammatical analysis of the sentence, and reduce the range of application of some of the rules. Such phrases have word functions in larger language units and fall into word-form classes. The Phrasifying Routine (Part I) (see Appendix V, Phrasification Flow) makes such treatment possible by accomplishing the recognition of the several word-records of the words in the phrase with a single word-record for the entire phrase. Recognition is done by comparison of a phrase entry prepared from input word-records with a list of phrase entries. If the constructed phrase entry is identical with one on the list, a phrase has been found. The list is prepared in advance by linguists. The routine, then, serves an entirely mechanical function, albeit somewhat complicated.

The mechanical aspect is complicated by two conditions of the problem: (1) the possible overlapping in the texts of phrases on the list and (2) the necessity of considering all combinations of words in the text not ruled out by certain practical and linguistic limitations from being possible phrases. The first condition necessitates a decision about what sequence of words is to be taken as the phrase in the various possible cases.

The general rule followed is (a) where the phrases begin with the same word, the longest and (b) where phrases overlap, the one which ends last is taken as the phrase. A practical limitation laid down is that phrases are no longer than twenty-four characters (including breaks between words). A linguistic limitation is that we do not look for phrases which include punctuation.

### Information Gathering

With data files prepared, updated and maintained in various sort sequences, the next stage in the data processing system is one of information gathering. A single program, the Affix-Dictionary program, has been written to gather information from the dictionary file and append it to the text file.

For the purposes of grammatical analysis, it is necessary to have preliminary assignments of words to form classes. Since the same word, however many times it occurs, will have the same preliminary word class assignments, it is most efficient to make this assignment by a mechanical dictionary look-up after the text word has been coded.

In addition, there are rules in English correlating regularities in word structures with correct form class assignments. Several words are often such that they can be thought of as "complex" words formed from one stem by the addition of prefixes and/or suffixes. These "complex" words often can be given a correct preliminary form class assignment on the basis of their affixes alone. This allows the number of words in a dictionary designed to provide word form-class information to be reduced radically. It is necessary to leave in only those words for which a preliminary word form-class assignment cannot be made on the basis of affixes alone. With

such a dictionary, if a dictionary lookup is performed and then affixes are noted and the rules assigning form-classes on the basis of them are applied, all words for which adequate provision has been made will receive their correct preliminary word form-class assignments.

It will, moreover, be useful to include the stems of the lawful "complex" words in the dictionary and to perform a lookup for them also provided there are items of information common to the words formed from a stem which are of use in grammatical analysis. The value of the rules of form-class assignment on the basis of affixes will be primarily the reduction in the length of the dictionary. In the mechanical table look-up, the length of the table is almost never a negligible consideration. Where a table the size of a dictionary is in question, any systematic reduction in size is bound to be important.

As it turns out there are items of information common to words whose form-class assignment depends upon affixes being added to the same stem. The reduction in dictionary length when only stems of these words are listed is still significant. So, a dictionary including incorrigibles and stems is used in the mechanical assignment of preliminary form-classes. The other dictionary information available in this way includes form subclass (see Appendix VI, Sub-Class codes) information which is not predictable from affixes. As would be expected from the above, which rule is applied to produce the preliminary form-class specification of a particular "complex" word depends upon the structure of the word. But it also turns out that if the additional dictionary information about the stem is utilized and the rules are made functions of items in that information,

the rules can be made more effective. This type of rule reduces the number of incorrigibles in the dictionary. Another pragmatic decision leads us to employ rules about suffixes only. We employ therefore, in our mechanical analysis of texts a Suffix-Dictionary routine for the production of dictionary information about words in a text.

### Affix-Dictionary Program

Input files to the Affix-Dictionary Program (see Appendix VIII, Affix-Dictionary Flow) are two: the dictionary file (Appendix II, File #1) and the text file (Appendix II, File #2). The program uses a binary look-up to avoid sorting and resorting. Since punctuation marks have been placed so that they have a zero in initial position, these are first. However, in these cases, dictionary information is zero and the main program is bypassed.

Output from the program is an appended text file (Appendix II, File #4) and an Error File (Appendix II, File #5). The Error File is produced when there is no matching dictionary entry for a word of the text or for its stem after affix removal. When this situation arises, additional entries must be made to the dictionary file by means of the Update Program. The Affix-Dictionary Program must then be repeated with the updated dictionary file as input.

By means of this program the 10-word record (Appendix II, File #2) representing a word of text increases to a 20-word record. The Affix-Dictionary Program inserts into the expanded record (Appendix II, File #4) codes representing all the form classes to which the word of text may belong, including a special code indicating a preferred form class, if any.

Subclassifications of form classes and another special code to indicate that an item is an "absolute breaker" (i.e. the item is the first word of a clause) are inserted. Coded representations of affixes either removed (before locating in the dictionary) or not removed (because the item was found) but indicated nevertheless, as well as the dictionary code which directs the program to test the ending of the dictionary word itself are included in the expanded record. Codes for irregular plurals, irregular past tenses, irregular superlatives and comparatives, past participles, and past forms which are always main verbs are also included (see Appendix VII, Dictionary Codes).

Although more information must be known about each text word before ambiguities can be resolved, the remaining information can be gathered in the Resolution of Ambiguities program which will be described in the next section.

Once the appended file has been produced it is sequenced into the analysable unit order (Text Identification) in order that the ambiguities may be resolved sequentially from left to right.

#### Resolution of Word Ambiguities

The resolution of word ambiguities within an analysable unit is carried out on the IBM 709 Computer in a manner which imitates the method employed by the linguist. The procedure in its simplest form might be stated as follows:

- 1) Within an analysable unit (left to right), note all words which are members of only one form class, i.e. never ambiguous;
- 2) Within the same unit (left to right), resolve the ambiguities of certain words of unusual distribution, as mine or like. (This step involves the application of one of a group of rules designed specifically for the

resolution of ambiguities of words in this class. Examination of the immediate environment of the words is required.)

3) Within the same unit (left to right), determine the particular type of ambiguity, such as noun/verb present tense or adjective/verb past tense. (The ambiguity is resolved by applying a suitable rule from a group designed for the resolution of this type of ambiguity. The application of the rule requires testing the immediate environment of the ambiguous word.)

The program written to follow this procedure consists of a control routine (See Appendix IX, Control Program) and a group of subroutines.

#### Control Routine

The control routine reads the analyzable unit from the appended text file into the computer memory, noting (step 1 above), as it reads, the words which are unambiguous. For each of these words the control routine places the appropriate English word, i.e., NOUN, into a specified part of its memory record.

When the unit has been completely read into the memory, the control routine begins its second pass. In this pass it determines the words belonging to the class defined as "words of idiosyncratic distribution" and transfers to a subroutine which applies rules sequentially until it resolves the ambiguity of the word. The subroutine then supplies the English word for the form class and the resolving rule number to the specified part of the memory record for this word of text. Control returns again to the control routine which proceeds in this fashion until it again reaches the end of the unit.

When all such words have been resolved, the control routine makes its third and final pass through the unit. In this pass it determines the particular type of ambiguity of all other ambiguous verbs and transfers them to the appropriate subroutines for resolving them. These subroutines function in the manner described in the preceding paragraph.

When all words of the unit have been resolved, the control program writes the unit whose records (Appendix II, File #6) now show the resolved form class and the resolving rule number.

The procedure is repeated until all the units of the text have been analyzed.

### Indicator Subroutine

Omitted from the above description of the control routine is reference to a special subroutine, the Indicator Subroutine. (Appendix X, Indicator Flow Chart). This subroutine completes the information gathering process begun in the Affix-Dictionary program and provides the control program and the remaining subroutines with sufficient facts for determining the type of ambiguity and for applying the specific rules for its resolution.

Resolution of form-class ambiguities depends upon the analysis of regularly recurring environments (indicator situations). First these indicator situations were broken down into parts; characteristics of the ambiguous item itself (for example, being the first word in the sentence, capitalization, etc.), characteristics of immediately preceding words, characteristics of preceding word + 1, characteristics of preceding word ignoring non-prepositional adverbs, and do on. Indicator categories were then set up and codes were given to each of these, indicator codes, which could be used in the machine. For example, one part of the indicator situation for several rules among the six rule set is the presence of a modal, copulative, or auxiliary verb. For one situation a member of this class is required to be immediately in front of the ambiguous item; for another, non-prepositional adverbs may be ignored. Again, for others, a member of this class must immediately follow the ambiguous item, in one case immediately following a word belonging to another category which itself follows the ambiguous item; in another adverb and adverb/adjectives may be ignored. Finally in one rule it

is only required to be the next verb. Nevertheless there is only one indicator code for this class - a "1" in the 10th position of the first indicator word. This means that each word in the sentence is tested to see if it is a member of this class, and if it is, a "1" is placed in the arbitrarily determined position of the arbitrarily chosen indicator word; otherwise there is a "0" in this position. Since this class is a subclass of a slightly larger indicator category a "1" would be put in the predetermined place indicating that the word is a member of this larger class also. Since both modals and auxiliaries are themselves members of other indicator categories, "indicators" must be placed in several places for such an item. Before any attempt is made to resolve the verbal ambiguities every word in the sentence must be tested with regard to these indicator classes (see Appendix XI, Indicator Codes).

On first consideration the task, performed by the control routine, of determining the type of ambiguity of a word seems relatively straightforward. It can be shown, however, to be quite complex, involving many computer instructions. For this reason the control program transfers to the Indicator Subroutine. The Indicator Subroutine performs the necessary tests and classifies each word by inserting a number in the Ambiguous Word Code (Appendix II, File #6). The control routine need then make only a single bit test to determine which type of ambiguity exists and, thereby, determine which subroutine must be entered.

In providing sufficient information for the subroutines to operate efficiently, the Indicator Subroutine is even more valuable. An illustration of a rule from one of the subroutines will serve to illustrate its value.

Is the preceding item to? If yes, does the ambiguous item have s as affix? If yes, take as noun. If no, see if the item before to is a nominal = the ambiguous item or according, alien, amenable, antagonistic, attributable, basic, complementary, contradictory, contrary, equivalent, foreign, hostile, inimical, liability, opposite, proportion(al), regard(s), resistant.



respect(s), sensitive, similar or supplementary, or if previous verb or adjective is part of ascribe, attack, attribute, belong, attain, cling, commit, convert, oppose, pertain, reconvert, relate, or subject. If yes, take as NOUN. If no, take as VERB.

Obviously the rule is complex even in the number of questions it asks before a resolution can be made. Each of the subroutines has between 20 and 40 rules of varying complexity. The additional complication of determining, for example, if the word in question is a member of one of the groups mentioned above increases the rule's complexity from a programming standpoint and decreases its flexibility.

Flexibility is a most important feature of this entire data processing system. It is especially important in the program for resolving ambiguities. There have been changes both in rules and rule ordering. There have also been insertions, deletions, and changes in word groupings such as the groups underlined in the above illustration. It is anticipated that when the results of the computer programs are studied, more changes are inevitable. A decision was made, therefore, that in order to maintain maximum flexibility the rules performed by the subroutines should be stated simply and directly in terms of computer instructions and that the determination of "belonging" to classes should be separated from the rules.

In the example above, the Indicator Subroutine "indicates" that a word of the unit does or does not belong to the group of words -- give, etc. by simply storing a "1" or "0", respectively, in a certain position of computer word in the memory record. The "1" and "0" are known as indicator codes, the computer word in the memory record is known as indicator code word. There are about 135 indicator codes. The groups which can be classified by means of indicator codes are varied. Among them are the following:

- 1) All words which are either object pronouns, or indefinite pronouns.

2) all words which are either possessive adjectives, or ambiguous precisely as adjective/pronouns, or ambiguous nouns.

3) all words which are either past participial adjectives or words which end in ing.

The control routine calls upon the Indicator Subroutine at several different times. By so doing, the information recorded in the indicator code words of the memory records is maintained in its most precise form for use by the subroutines. Instead of testing a word against a long list, for example, to determine whether or not a rule is suitable for resolving a word's ambiguity, the subroutine need test only a single bit position of an indicator code word.

#### Other Subroutines

It is by means of the subroutines that the computer resolves the ambiguous words of the text. There are six subroutines which correspond to the six types of ambiguities which are to be resolved. Each subroutine is made up of a) a control program, b) a rule table, and c) coding which represents, in computer terms, each of the rules in the set for resolving the specific ambiguity.

The subroutine control program is standard for all the subroutines. It may be stated as follows:

- 1) Advance a rule counter C.
- 2) Locate from a rule table the starting address of the coding corresponding to Rule C.
- 3) Transfer to the coding for Rule C.

The coding for Rule C either resolves the ambiguity or returns to step 1) above. If the ambiguity is resolved, the resolved part of speech

and the resolving rule number C is supplied to the specified portion of the text word in memory. The counter is reset to zero and the subroutine returns to the main Control Routine.

The rule table consists of a list of symbolic addresses corresponding to the entry points to coding for each rule in the set. The subroutine executes the rules by transferring in sequence to the symbolic addresses in this table.

The use of a rule table adds another feature of flexibility to the system. By rearranging the sequence of the symbolic addresses in the table it is possible to rearrange the order in which the rules of the set are applied. It is further possible to eliminate the application of certain rules by simply omitting their symbolic addresses from the table. Rules may also be added by supplying the necessary coding for the rule, assigning it a symbolic address, and inserting this address in the desired place in the table. These changes are made by reassembling the program. The rule number is not actually attached to any piece of coding, but corresponds to the order of rule application.

The coding for a rule in the set may require the examination of the indicator codes, the form classes, and other characteristics of the ambiguous word or of the words which precede or follow it. When examination of words other than the ambiguous word is required, the subroutine gains access to these words by means of special "search" routines.

The search routines (see APPENDIX XII - The Search Routines) were developed to meet the common need of all the subroutines. There are eight search routines: four to locate words preceding the given word and four to locate words following.

### The Absolute Breaker Subroutine

In addition to absolute breaker items which always begin a clause, there are simple situations, involving only two or three words, which always indicate that at least one of the items begins a clause. These situations normally indicate that any search must stop there. They may also serve to limit the possibilities for an ambiguous item which precedes one of them. This program has been written to place the breaks as indicated in the appendix (see APPENDIX IV, Absolute Breaker Lists).

Construction of a Routine to Mark Clause Divisions in a Partially  
Analyzed Sentence.

A syntactical analysis was defined as an ordered sequence of syntactical symbols. The device to produce syntactical analysis for any English sentence, which we are describing, treats the processes of assigning syntactic symbols to words and of ordering these symbols as being interdependent. Sufficient information is derived for a string of symbols produced by the look-up routine to allow a partial ordering into clauses to be imposed upon it. This partially ordered string in its turn contains sufficient information to enable either a proper assignment of symbols to be made for all words and ordering into phrases within the clauses to be carried out, or corrections in previous wrong assignments of syntactic symbols to be made resulting in a re-ordering into clauses, in which items improperly labeled can then be assigned an ambiguous symbol and ordering into phrases completed.

Division into clauses is carried out in many cases on improper strings of symbols, i.e. strings containing ambiguous symbols. Since the ambiguity routines described above have already been called in there are no improper strings containing the element verb. A search is initiated at the end of the string and continued forward to the beginning until either a verb or an absolute breaker is found. An absolute breaker is a word like which, what, because or a phrase like just as though, in order to which regularly stands at the beginning of a clause. These words and phrases are contained in the dictionary and are marked as absolute breakers. The members of another set of absolute breakers are identified by a search routine. They are either juxtapositions of items which are impossible within the same clause such as Verb + Subject Pronoun or items which can occur together within the same phrase

and whose juxtaposition indicates that the first element is joining two clauses, such as that + definite or indefinite article. Present participles, infinitives and past participles ending in -en like given, written, etc., when not preceded by auxiliary verbs are also absolute breakers. Lists of absolute breakers currently recognized by the routine are given in Appendix O. The beginning of a sentence counts as an absolute breaker.

In the case where the search finds an absolute breaker, the beginning of a clause is marked as occurring there and a new search is initiated at this point. In the case where the search finds a verb before coming to an absolute breaker this fact is registered. If, after this, an absolute breaker is found before another verb, a clause division is marked at the absolute breaker and a new search starts. If another verb is found before an absolute breaker the situation becomes somewhat more complicated. It is certain that somewhere between the two verbs a clause division has been crossed since no clause can contain more than one verb (unless one of them is an auxiliary or a modal).

In order to discover where this division lies it is necessary in most cases to examine the characteristics of the two verbs involved. For this it is necessary to make use of the sub-class information contained in the dictionary. Particularly relevant is the sub-classification of verbs; transitive, intransitive and transitive/intransitive. The term transitive here is given a special interpretation. It is applied not to verbs which must always take an object (i.e. must be followed by a nominal group not headed by a preposition) but to those which always do so when they are the main verb of a sentence. The verb fill for example is marked in the dictionary as transitive, since it will only occur without an object when it occurs

as a passive participle or in a passive phrase (Cf. He filled the sack with flour. He lifted the sack filled with flour. The sack was filled with flour.).

This provides the basis for a rule for determining clause boundaries.

In a situation in which a clause boundary has to be established between two verbs (what we shall call from here a redistribution situation) if the right hand verb is transitive and has an -ed suffix, and is not followed by an object then the clause boundary is to be drawn immediately in front of it as in He lifted the sack/filled with flour. It doesn't matter which sub-class the left hand verb belongs to; nor what tense it is. In these cases it will always form part of a degenerate clause, functioning as a post-nominal modifier of the preceding noun. The importance of the sub-classification transitive and intransitive provides the rationale for beginning the clause division routine at the end of the sentence. Once the end of a clause has been established (and in nearly all cases the end of the last clause is the end of the sentence), then to discover whether the verb in that clause has an object is a comparatively simple procedure. It is only necessary to establish that somewhere before the end of the clause there occurs a noun or some element of a nominal group such as a definite or indefinite article, or adjective, or a demonstrative pronoun, or an object pronoun, etc. which does not have a preposition immediately in front of it.

Another important sub-class as far as its usefulness in clause determination is concerned, is that to which all verbs that can take a full clause as their object belong. Examples include verbs like believe, hope and notice (I believe that he will come tomorrow. I hope she is not ill.). The first of these examples suggests another rule (though in point of fact both that + he and hope + she - (verb + subject pronoun) - are absolute breakers so there is

no redistribution situation in these sentences). If in a redistribution situation the left hand verb belongs to this sub-class and somewhere between the two verbs there occurs the item that, the clause break occurs there I noticed/ that one of the students was asleep.

Certain important points arise out of this example. The only information we need about the left hand verb is that it belongs to the sub-class in question. It does not matter to what other sub-classes it also belongs. (Few English verbs fit into only one sub-class. Any attempt to arrange English verbs - or nouns - into classes in such a way that the maximum amount of information about their behavior is utilized results in considerable cross-classification. Schemes indicating the combinations of sub-class membership of verbs and nouns now recognized by the look-up routine will be found in Appendix F). As long as the look-up routine reveals that the verb in question can function in this way the rule applies.

But notice that in the case of a sentence like I noticed that sack filled with flour our present rule conflicts with the previous one. It produces a wrong answer whereas the previous rule would produce a right answer. As was the case with the ambiguity rules the rules for clause division must be ordered. It is essential that our first rule is applied before this one.

A correct solution can often be established on the evidence of tense alone without any use of sub-class information. Take the sentence The man opening the door works in my office. In any redistribution situation where the left hand verb is a present participle and the right hand verb is present tense the break must come immediately in front of the right hand verb.



Notice that since a present participle is also an absolute breaker the final ordering will be The man/opening the door/works in my office. In these cases provision must be made associating the two parts of the interrupted clause.

In certain redistribution situations a solution can be reached without reference to the verbs at all. The simplest case of this kind occurs when the two verbs are only separated by adverbs or prepositions, as in, He dined well/read a book/ and went to bed early (note and + verb is an absolute breaker). Notice that in this case the rule must stipulate that only these items occur between the verbs. The occurrence of a noun, for example, would produce wrong answers when the left hand verb belonged to the sub-class of verbs which can take a clause as their object.<sup>12</sup> In order to keep the rules as general as possible many of them refer to items whose presence makes the rule inapplicable. In each case the situation produced by the presence of such an item is resolved by a later, and usually a general, rule.

Another redistribution situation which can be analyzed without reference to the characteristics of the verbs involved is illustrated in the sentence, The thought that he had been treated in such a cruel fashion shocked . . . . It will be noticed that if no nominal group that can be a subject occurs between the two verbs (this is marked by the same criteria as object, i.e. a noun or element of a nominal group must be found not preceded by a preposition)

---

<sup>12</sup> There is a problem here with certain time and space expressions which are formed like nominal groups but structure like adverbs, e.g. Drink a pint of milk a day, He came last summer. He went home. It is hoped that this problem can be overcome by listing these phrases in the dictionary with the assignment Noun/Adverb. A word or phrase with this assignment would not afford an exception to this case.

then the cut must come immediately in front of the right hand verb. The sentence The thought he had treated his children in such a cruel fashion shocked them shows that the rule can be extended by making the condition that there should be only one nominal group after the first preposition found searching from right to left. The rule can be still further extended by making the more specific restriction that the nominal group after the preposition must not contain two or more nouns if the second is either a plural noun or an uncountable noun (by which is meant a noun which can occur immediately in front of a verb without either an article, demonstrative adjective or possessive adjective before it, or a plural suffix e.g. milk, envy). The sentence The thought his guests had behaved in such a cavalier manner shocked him will now be correctly handled by this rule. The new restriction prevents the second noun being taken as a separate nominal group. In the case of a sentence like He talked in a manner men tended to dislike since the second noun after the preposition is plural the rule is inapplicable.

This kind of situation is dealt with by another rule which makes no use of verbal information. This rule lists all the juxtapositions of elements indicating the coincidence of two nominal groups and commands that a clause boundary be marked between them. The sequence indefinite article + singular noun + plural noun which occurs in the sentence above is just one of many examples. Others include noun + subject pronoun - He saw the book he wanted, noun + definite or indefinite article - He saw the book the girl wanted, plural noun + noun - He saw the books John wanted. Notice that these combinations cannot be made absolute breakers because they only mark the occurrence of a clause division when they occur in a redistribution situation. In cases like In the garden the flowers were blooming there is, of course, no clause break between the noun and the article. An absolute breaker, on the other hand always

indicates a clause division. Notice also that none of these combinations can be taken as diagnostic if the left hand verb belongs to the sub-class of object verbs (give, bring, elect, etc.) and the right hand verb is transitive and has an -ed suffix. In all these cases there seems to be an irreducible syntactical ambiguity. Compare:

He showed the group the operators/selected  
and  
He showed the group/the operators selected.

Here one needs a rule which states that under these circumstances two analyses must be provided; though it is still doubtful whether this distinction is necessary for information-retrieval purposes.

In some cases it is necessary to look in front of the redistribution situation. If the left hand verb has certain absolute breakers such as who or which immediately in front of it then the break comes immediately in front of the right hand verb. The woman who bought this house last summer has recently married. This rule must come after the rule discussed in the preceding paragraph if the routine is to handle correctly sentences like The woman, who patted the dog the little boy had with him, was very fond of animals.

So far all the redistribution situations we have discussed arise when a clause division is sought between two verbs. This is a result of the decision to adopt the operational definition of a clause as that part of a sentence containing one and only one finite verb. There are, of course, clauses which do not contain verbs. (These are transforms by deletion of full clauses.) Examples would be, The professor, while still a young man, had studied in many universities. The figures though inexact were still useful. To deal with these cases we add to the redistribution rules two special rules applied before all others. The first states that if between an absolute breaker and a verb, there occurs one and only one comma then a clause division is to be marked as occurring at that comma. The second states that if between an absolute breaker and a verb

there occurs an adjective but no noun then a break occurs before the right verb. Probably more rules will need to be added to this group.

The rules are structured in such a way as to make them as flexible as possible. Each contains the minimum of information needed to establish the existence of a context in which a particular decision can be made. The only other information given is a list of those elements whose presence changes the context. A schematic representation of the rules is presented in appendix P. The list is not yet complete. There are two outstanding omissions. First, in the present rules no account is taken of the special circumstances arising when the left verb is a two object verb except to mark those cases where its occurrence makes one of the present rules inapplicable. Second, the rules presented here are not designed to cover those cases where clauses are joined together by conjunctions like and or but. Considerable problems arise in these cases since these conjunctions can join either two words or two clauses. Cf. I like coffee and John likes tea and I like coffee and tea made with milk. Most of these have been solved and rules that establish which function the ~~conjunction~~ is fulfilling (and when it can fulfil both as in I noticed the boy and the girl saw me) drawn up. This work however, is not yet ready for publication.

### III

#### 1. FLIX: A Propositional Form for Informational Retrieval

The basic structure of FLIX is represented in the following diagram.

---

$S_1$	$S_2$	$S_3 \dots$	$P_1$	$P_2$	$P_3 \dots$
Subject	Qualifiers of $S_1$	Qualifiers of $S_2$	Verb or predicated noun or adjective	Qualifiers of $P_1$	Qualifiers of $P_2$

---

Consider a standard sentence, which can take one of the following forms

---

$S_1$	$S_2$	$S_3$	$P_1$	$P_2$	$P_3$	$P_4$
			Verb	Noun (obj.)	Adjective	Adverb
				Adverb		
Noun	Adjective	Adverb	Noun (pred.)	Adjective	Adverb	
			Adjective (pred.)	Adverb		

---

Note, in passing, that the scheme as shown is incomplete, since an adverb can be qualified by an adverb, which may in turn be qualified by a further adverb. This is accommodated in the natural way by entering, as usual, the new qualifier in the column to the right of the word which it qualifies. Thus the following sentence is essentially of the standard form:  
 "A quite exceptionally drunk man offered a violent affront to a really stupendously slowly moving onlooker." and is written:

$S_1$	$S_2$	$S_3$	$S_4$	$P_1$	$P_2$	$P_3$	$P_4$	$P_5$	$P_6$
Man	drunk	exceptionally	quite	offered	affront	violent	moving	slowly	stupendously
					onlooker				

---

$P_7$   
really

---

We now deal with clauses and phrases, according to the following rules:

For a given clause ascertain the FLEX rating of the word for which it substitutes. Thus, in the sentence "to be quite exceptionally drunk offers an affront, which can scarcely be overlooked, to what we are pleased to call the moral fabric of our way of life," the clause "to be quite exceptionally drunk" substitutes for a noun standing as subject of the sentence. This noun would have been rated  $S_1$ . A second clause "which can scarcely be overlooked" standings in the same relation to "affront" as does the adjective violent rated  $P_3$  in the earlier sentence. Finally, the clause "what we are pleased to call the moral fabric of our way of life" substitutes for a noun (cf. "onlooker" in the earlier sentence) which would receive the rating  $P_2$ .

Representing the FLEX rating of a word as  $X_n$ , where X stands for S or P and n stands for the numerical suffix, enter the clause according to the following format:

---

$X_n$	$X_n + 1$	$X_n + 2$	$X_n + 3$	$X_n + 4$
Noun (subj.)	Adjective	Adverb		
Verb	Noun (obj.)	Adjective	Adverb	
	Adverb			
Noun (pred.)	Adjective	Adverb		
Adjective (pred.)	Adverb			

---

Comparison with the standard arrangement on page 10 will make clear that the leading words in the clause (i.e. those which would be rated  $S_1$  or  $P_1$  if the clause were a main clause) have simply been given the rating due to the word for which the clause substitutes, and that the remaining words have been anchored to these in precisely the same fashion as was done for the standard sentence.

The treatment of phrases is simpler, the formula being

$X_n$	$X_n + 1$	$X_n + 2$
Noun	Adjective	Adverb

The sentence given earlier appears in FLEX as follows:

---

$S_1$	$S_2$	$S_3$	$P_1$	$P_2$	$P_3$	$P_4$	$P_5$	$P_6$
drunk	exceptionally	quite	offers	affront	overlooked	scarcely	moral	our
					call	pleased	way	life
						fabric		

---

The procedure as described is sufficiently general to extend in an obvious fashion to clauses within clauses, phrases within clauses within phrases, etc.

Determination of the relevance of a store sentence to the input sentence.

This is based on word-by-word matching. The goodness of match between a store word and an input word is treated as a function of two variables:

- (1) similarity of FLEX rating.
- (2) semantic proximity.

To illustrate the use of criterion (1), consider the two sentences

A. Socrates as a young man ran to Athens.

B. A young man from Athens ran Socrates into debt.

The words belonging to the major form classes, i.e. those to be subjected to the matching process, have been italicized. In five cases out of six there is perfect matching both for form class and for semantic content. Yet the mutual relevance of the two sentences is practically nil, and should become clear as soon as any kind of linguistic analysis is applied.

The FLEX versions are as follows:

S <sub>1</sub>	S <sub>2</sub>	S <sub>3</sub>	P <sub>1</sub>	P <sub>2</sub>
A. Socrates	man	young	ran	Athens
B. Man	young		ran	Socrates
	Athens			debt

Successful use of FLEX depends upon the numerical system whereby the FLEX ratings of the two words being matched are made to contribute to the overall score for goodness of match. The system gives weight to the importance of the words in relation to the rest of their sentences: a perfect match between words which are making an unimportant contribution to their sentences will in general have less bearing on sentence-relevance than an imperfect



match between two important words, e.g. the two subjects, or the two main verbs. FLEX gives a rough guide, the categories being ranked for importance as follows:

$S_1 \quad S_2 \quad S_3 \dots\dots\dots$  and  $P_1 \quad P_2 \quad P_3 \dots\dots\dots$

The system we propose gives due weight not only to importance in itself but also to degree of cross-correspondence between the FLEX ratings of the two words. This is shown in detail in a later section.

### Semantic matching.

Our basis of semantic matching is the use of a thesaurus. By a thesaurus we simply mean a list of clusters of words, each cluster bearing an index number. Words are assigned to the same cluster if their mutual semantic relationship exceeds some threshold. A given word may, and indeed often does, appear in many different clusters, reflecting the variety of shades of meaning which one and the same word may bear when used in different contexts. An extreme example might be the word induction which has technical meanings in such diverse fields as electromagnetism, biology, and logic, in addition to a range of more loosely defined connotations when used in ordinary speech. Ideally the next phase of the project should include the construction of a thesaurus suitable for modern information-retrieval. Proposed methods for constructing ab initio a more suitable thesaurus will be explained in a later section. For testing the numerical procedures under discussion we have been content to use Roget's Thesaurus.

The basis of our approach to semantic matching is the principle that the more closely related semantically two words are, and the more frequent the occasions on which they can be used interchangeably, the greater the number of clusters which they will have in common in any rationally constructed thesaurus. Thus, the word diet has three clusters listed in Roget's index,

with index numbers 298, 662, and 696, and the word nutritional has four clusters with index numbers 298, 662, 656, and 707. (In using the thesaurus we treat all derivatives of the same root, e.g. nutritional, nutrition, nutritious, nutritive, nutrient, etc. as being variants of one and the same word.) There are two clusters in common, indicating a rather high degree of overlap in view of the fewness of the total number of clusters involved. A serviceable measure of semantic correlation based on cluster overlap, is given by the formula

$$\frac{n_{ab}}{\sqrt{n_a \times n_b}}$$

where  $n_{ab}$  is the number of clusters common to both words,  $n_a$  is the number of clusters indexed under word a and  $n_b$  is the number under word b. The total number of clusters involved is automatically allowed for. Applying the expression to the example just given, we have

$$\frac{2}{\sqrt{3 \times 4}} = 0.58.$$

Note that the highest value attainable is 1.00, indicating that the two words share all their clusters, and that the lowest value is zero.

#### Measuring mutual relevance of two sentences on semantic criteria alone.

We are now in a position to illustrate the use of the above measure in an actual comparison of sentences in a stored text with the input sentence. We suppose that there are present in the stored text the following two sentences:

1. Vegetables is suitably prepared can provide all the nutritional constituents necessary for growth.
2. It is essential to ensure transport facilities for the supply of components to the installation.

Let us consider the following input sentence:

Do plants supply the essential components of a balanced diet?

The FLEXI versions of these three sentences are given below.

---

	$S_1$	$S_2$	$S_3$	$P_1$	$P_2$	$P_3$	$P_4$	$P_5$
1.	Vegetables	prepared	suitably	provide	constituents	nutritional necessary	growth	
2.				ensure	facilities	transport supplying	components	instal- lation
Input:	Plants			supply	components	essential diet	balanced	

---

We now prepare a sentence correlation table, preserving the FLEXI format, as in table 1.

		S <sub>1</sub> plants	P <sub>1</sub> supply	P <sub>2</sub> components	P <sub>3</sub> essential	P <sub>3</sub> diet	P <sub>4</sub> balanced
S <sub>1</sub>	vegetables	0.17	0	0	0	0	0
S <sub>2</sub>	prepared	0	0	0.09	0	0	0
S <sub>3</sub>	suitably	0	0	0	0	0	0
P <sub>1</sub>	provide	0	0.20	0.06	0	0.14	0
P <sub>2</sub>	constituents	0	0	0.23	0.13	0	0
P <sub>3</sub>	nutritional	0	0.20	0	0	0.58	0
P <sub>3</sub>	necessary	0	0	0	0.17	0	0
P <sub>4</sub>	growth	0.15	0	0.09	0	0	0
P <sub>1</sub>	ensure	0	0	0	0	0	0
P <sub>2</sub>	facilities	0	0	0	0	0	0.14
P <sub>3</sub>	transport	0	0	0	0	0	0
P <sub>3</sub>	supplying	0	1.00	0.09	0	0	0.12
P <sub>4</sub>	components	0	0.09	1.00	0	0	0
P <sub>5</sub>	installation	0.15	0	0	0	0	0

Table 1. Upper sentence, semantic correlation =  $\frac{0.17 + 0.20 + \dots}{\sqrt{8 \times 6}} = 0.32$

Lower sentence, semantic correlation =  $\frac{0.14 + 1.00 + \dots}{\sqrt{6 \times 6}} = 0.43$

The semantic correlations between the words of the input sentence and the words of the store sentence, calculated according to the formula previously given, are entered in the corresponding cells of the table. It will be noted that the sum of the individual correlations is scaled down in the final expression, by a factor derived from the total number of words involved, thus allowing for the influence of sentence length on the expected number of matches.

This example was concocted in order to illustrate the nature of the limitations of a retrieval method based purely on semantic correlation. Sentence 2 of the stored text was deliberately designed to show heavy semantic overlap with words of the input sentence, in spite of bearing little or no relevance to it. This is reflected in the final score of 0.43 as compared with 0.32. Using semantic criteria alone, the wrong sentence would be retrieved. Information retrieval systems which depend purely on semantic matching without recourse to syntactic analysis, have to accept an irreducible load of such errors as a basic limitation of their approach. The use of FLEX as a simplified syntactic system offers at least a partial remedy for this shortcoming, as will now be illustrated by the introduction of FLEX correlation into the foregoing worked example.

#### Numerical use of FLEX for measuring relevance

Before proceeding to detailed illustration, one or two comments on points of detail are called for.

(1) It may seem surprising that in addition to the disappearance from the FLEX format of subsidiary words such as prepositions, connectives, etc., the entire clause it is essential has vanished from sentence 2, and the noun clause, of which this was the predicate, now appears in the predicate section of the FLEX format. This is a consequence of one of several sophistications

with which FLEX has recently been endowed, following from the discovery that certain syntactic patterns mark "dummy" constructions which are regularly used as equivalent to an active statement with absent or vaguely defined subjects.

(2) The FLEX version of sentence 2 has wrongly assigned the word installation as a modifier of components instead of supplying. This arises from an imprecision in our current routines for assigning word groups which we hope to remedy.

In constructing numerical measures for the goodness of match between one FLEX category and another we have been guided by two principles, both of which must receive due expression in the final measure of correlation. They are,

(1) The relative importance of the category. Words appearing, for example, in  $S_1$  or  $P_1$  are likely to contribute a much greater weight to the relevance of the sentence to some other sentence than words appearing, for example, in  $P_5$ .

(2) The degree of correspondence between the categories of the two words which are being compared. Thus if we are comparing a word in  $P_1$  with a word in  $S_1$ , the final score for the match should plainly be less than that derived from a comparison between a word in  $P_1$  with a word in  $P_1$ , or from an  $S_1 - S_2$  match. In the same way, let us contrast a  $P_1 - P_3$  match with a  $P_2 - P_2$  match: although the average level of importance is the same in the two cases, it is obvious that the latter should receive the higher score.

The numerical values at which we finally arrived are those set out in table 2.

$s_1$	1.00	0.63	0.34	0.18	0.09	0.71	0.47	etc.		
$s_2$	0.63	0.50	0.32	0.17	0.09	0.47	etc.			
$s_3$	0.34	0.32	0.25	0.16	0.09	0.28	etc.			
$s_4$	0.18	0.17	0.16	0.13	0.08	0.16	etc.			
$s_5$	0.09	0.09	0.09	0.08	0.06	0.08	etc.			
$P_1$	0.71	0.47	0.28	0.16	0.08	1.00	etc.			
$P_2$	0.47	0.35	0.24	0.14	0.08	0.63	0.50	etc.		
$P_3$	0.28	0.24	0.18	0.12	0.07	0.34	0.32	0.25	etc.	
$P_4$	0.16	0.14	0.12	0.09	0.06	0.18	0.17	0.16	0.13	etc.
$P_5$	0.08	0.08	0.07	0.06	0.04	0.09	0.09	0.09	0.08	0.06
	$s_1$	$s_2$	$s_3$	$s_4$	$s_5$	$P_1$	$P_2$	$P_3$	$P_4$	$P_5$

Table 2. FLEX correlations (Note that the upper right corner duplicates the lower left, and lower right equals upper left.)

They were derived from a geometric model which is appended in figure 1.

The quantities shown in table 2 are applied as multipliers in the correlation table, as shown in table 3. The products are summed and converted into revised scores. It will be observed that the relative ordering of the two sentences in the stored text with respect to relevance has been reversed. In other words, the high level of fortuitous semantic overlap between the input sentence and the "wrong" sentence of the stored text has been efficiently counteracted by the use of FLEX correlation. We have not yet performed the extensive tests on specimen passages of stored text which will be necessary before we can say just how effective and versatile the method will prove to be, but the preliminary indications have been highly encouraging. Various improvements have already suggested themselves, but it would take us too far into technicalities to detail them here. (See Appendix XIX - Concerning the Matching Formulae).



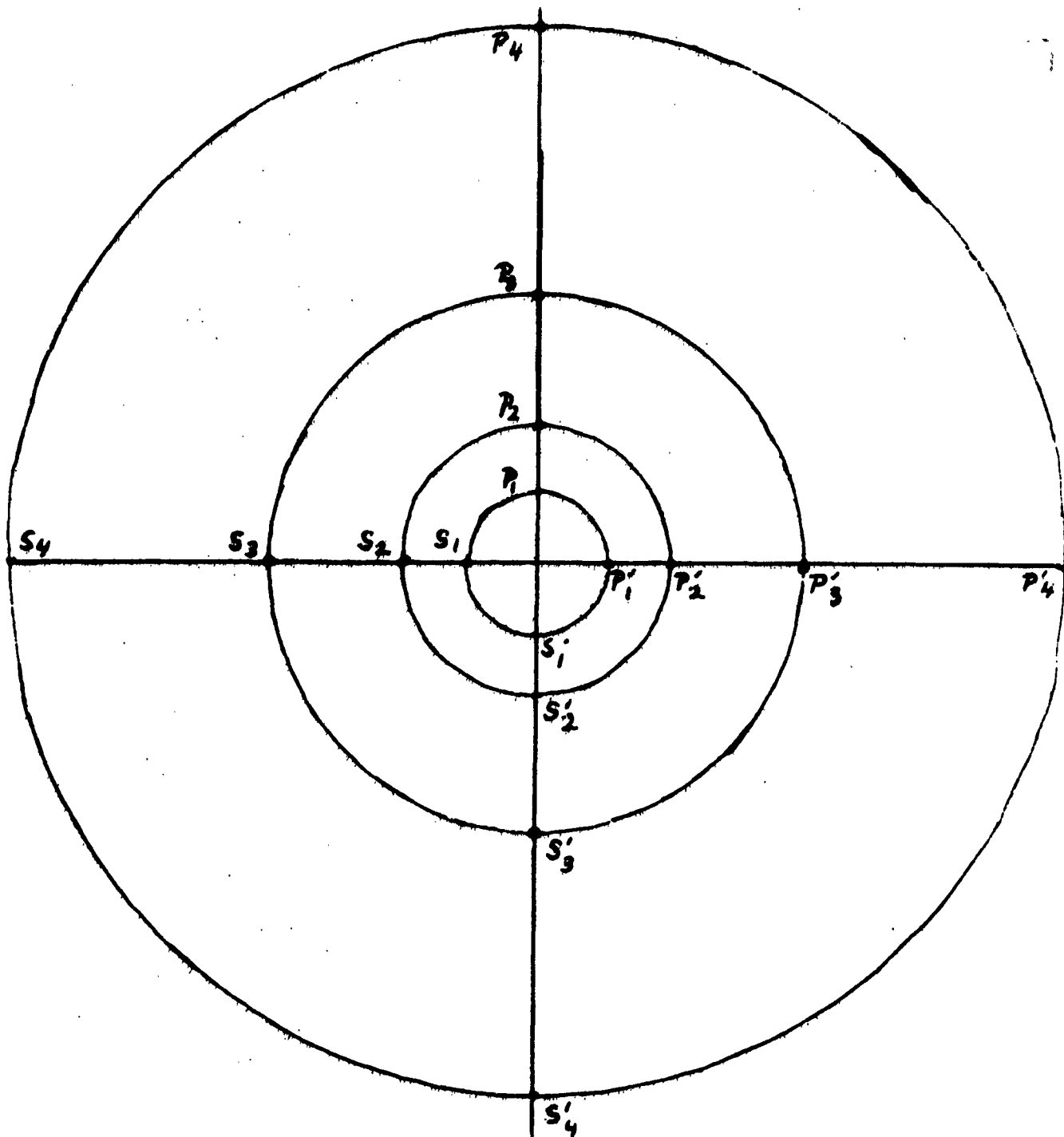


Fig. 1. Geometric model from which the FLEX correlations were constructed.

$S_1, P_1$  etc. represent words of stored text, and  $S'_1, P'_1$  etc. represent input words. The radii of successive circles are  $\frac{1}{\sqrt{2}} \{1, 2, 4, 8, 16 \dots \text{etc.}\}$ . The FLEX correlation is given by the reciprocal of the distance between any two points.

	S <sub>1</sub> plants	P <sub>1</sub> supply	P <sub>2</sub> components	P <sub>3</sub> essential	P <sub>3</sub> diet	P <sub>4</sub> balanced
S <sub>1</sub> vegetables	0.17x1.00 = 0.17	0	0	0	0	0
S <sub>2</sub> prepared	0	0	0.09x0.35 = 0.03	0	0	0
S <sub>3</sub> suitably	0	0	0	0	0	0
P <sub>1</sub> provide	0	0.20x1.00 = 0.20	0.06x0.63 = 0.04	0	0.14x0.34 = 0.05	0
P <sub>2</sub> constituents	0	0	0.23x0.50 = 0.11	0.13x0.32 = 0.04	0	0
P <sub>3</sub> nutritional	0	0.20x0.34 = 0.07	0	0	0.58x0.25 = 0.15	0
P <sub>3</sub> necessary	0	0	0	0.17x0.25 = 0.04	0	0
P <sub>4</sub> growth	0.15x0.16 = 0.02	0	0.09x0.17 = 0.02	0	0	0
P <sub>1</sub> ensure	0	0	0	0	0	0
P <sub>2</sub> facilities	0	0	0	0	0	0.14x0.17 = 0.02
P <sub>3</sub> transport	0	0	0	0	0	0
P <sub>3</sub> supplying	0	1.00x0.34 = 0.34	0.09x0.32 = 0.03	0	0	0.12x0.16 = 0.02
P <sub>4</sub> components	0	0.09x0.18 = 0.02	1.00x0.17 = 0.17	0	0	0
P <sub>5</sub> installation	0.15x0.08 = 0.01	0	0	0	0	0

Table 3. Upper sentence, composite correlation =  $\frac{0.17+0.03+\dots}{\sqrt{8 \times 6}} = 0.14$

Lower sentence, composite correlation =  $\frac{0.02+0.34+\dots}{\sqrt{6 \times 6}} = 0.10$

Fully mechanized construction of a thesaurus suitable for numerical methods.

Existing scholarly compilations such as Roget's are unsuitable on grounds of vocabulary for mechanized information retrieval. They also lack any consistent quantitative basis for the assignment of words to clusters, or for the delimitation of cluster boundaries. More objective methods exist whereby clusters could in principle be constructed on the basis of a common underlying scale of measurement. These, however, take as their starting point our ability to give semantic proximities between pairs of words at least a rough-and-ready ranking order. Since the meaning of words can only be defined in terms of their significance to human beings, it would seem that we are forced back upon our own subjective judgment, and that the mechanization of thesaurus-making is in principle unattainable. The enormous labor of making and arranging judgments of this type has however already been done by generations of lexicographers, and there is no reason why the text of an existing large dictionary should not be used as input for a thesaurus-making computer program. We have made preliminary hand tests of an extremely simple procedure for ordering the semantic proximities between pairs of words, using Webster's Third New International Dictionary. The method is again based on overlaps, this time on the degree of overlap between the words occurring in the definitions of the two words concerned. It has surprised us to find that the results seem as reliable as those achieved by the subjective judgment of professional linguists. We envisage that the fully mechanized procedure would require the following steps:

- (1) The edited text of a suitable dictionary to be punched on cards,
- (2) A computer program to be written which computes the overlapping between words taken pairwise,

(3) A computer program to be written, or an existing one\* utilized, which uses the measures obtained by (2) to pack the words into a "semantic space" of minimum dimensionality and to assign them coordinates within this space,

(4) A computer program to be written which will "sweep" the semantic space, gathering the words up into clusters.

In order to reduce the above procedures to easily manageable proportions it would be necessary to make a prior coarse grouping of words into superclusters using an existing thesaurus. Our main retrieval project is in no way critically dependent on the preparation of such an ideal thesaurus; the proposed work can go ahead on the basis of a compilation such as Roget's until a refined and improved version becomes available to replace it.

---

\* R. N. Shepard of Bell Telephone Laboratories has written a program designed for other purposes which could readily be adapted to our needs.

## 2. Outline of Information Retrieval Process

### A. Four tables are used.

(1) Dictionary with thesaurus-concordance. This table consists of a list of word stems and phrases. Against each item is entered a string of numbers referring to the clusters in the thesaurus in which the word appears. We refer to these numbers as thesaurus references.

(2) Thesaurus with text-concordance. This table consists of a numbered list of clusters. Against each cluster is entered a string of numbers referring to the text in the corpus containing a word of the cluster having high absolute information content. (See Appendix XIX - Concerning the Matching Formulae.)

(3) Text to tape concordance.

(4) Table of FLEX correlations (as in Table 2).

B. The corpus from which it is desired to retrieve information has its FLEX assignments and thesaurus references in each item record.

C. The machine assigns the proper FLEX labels to the items of the question sentences.

D. The corpus sentences are brought in and matched with the question sentence.

E. The answer paragraphs with their identification numbers are printed out in the order of their matching scores.

The sentence matching procedure could be as follows.

Each question sentence item undergoes whatever affix stripping is necessary to find a stem (with the same thesaurus references) in the dictionary. Information content, the number of thesaurus references, and the thesaurus references are noted for each word.

A composite list of thesaurus entries is made in thesaurus order. From this list, the pertinent texts are listed in corpus order. (If necessary the needed tape numbers could be printed out for the librarian to put on the tape

readers.) The text sentences are then matched word for word with the input question and a list of paragraphs, their identification numbers, and the corresponding sentence matching numbers then generated.

This list of paragraphs can then be ordered by means of their highest sentence matching numbers for printing out.

#### IV. Conclusions and Recommendations

a. On the utility of logically structured languages. Although this portion of the investigation was never pushed intensively, enough was done to suggest the following: (1) that coding information in an artificial language suitable for logical manipulations is suitable for certain special kinds of data only, and not for the main mass of information in most relatively discursive fields; (2) that in the very restricted areas where it is suitable, further investigation is warranted.

b. On complete tape storage of all data. Here it is obvious that present hardware is inadequate to permit fast access to information which is merely known to be somewhere in a tape file containing material in the order of billions of words of text. If the hardware problem can be solved (e.g. by improved photoscopic discs or the like), it may become economical to develop searching techniques such as those presented in this report.

c. On machine "learning" programs for syntax. Ellson's experiment shows conclusively that these cannot be made to reach a useful level of accuracy if the only form-class information available is that given in standard published monolingual dictionaries of English. However, if dictionaries containing a much more elaborate form-class breakdown (into 100 or more subclasses instead of eight or ten) ever become available, it will probably be well worth while to redo the experiment.

d. On semantic digital coding of vocabulary. Here again the research was insufficient for firm conclusions, but it does appear that considerable economies of storage might be achieved by such means if translation programs could ever be rendered fully automatic.

e. On the value of a randomly selected corpus for testing. Here only guesses are available; a priori such a corpus ought to show up any deficiencies in MT or IR programs, but we have no evidence to prove this.

f. On the key-punching bottleneck. It is obvious that no efficient machine program is possible without a really fast and accurate print-reading device.

g. On dictionary codes. One of the most immediate and urgent needs is a large fully coded dictionary, i.e. one in which every potentially useful fact of syntax and co-occurrence possibilities is indicated for every word, except those for which the full range may automatically be determined from a consideration of the affixes present.

h. On affix analysis. As implied in the preceding paragraph, affix analysis routines have at least two strong utilities, both of which should ultimately be used in any efficient IR program: (1) for reduction of the dictionary, both in total entries and in need for separate hand coding; (2) for automatic association of semantically linked words. The present project has not fully exploited either of these as yet. Our reverse-alphabetized dictionary and all other present and future reverse-alphabetized dictionaries will be of essential utility in further work on such programs.

i. On FLEX. Here, again, we have only begun. So far, it appears (1) that elements which are syntactically marked as heads are generally also of more informational importance, and (2) that the grammatical subject has a distinctly different informational role than the grammatical predicate. It is possible, however, that for some types of transitive verbs (including phrase-verbs consisting of intransitive verb plus prepositions the object is informationally indistinguishable from a subject. For example, the sentence "We enjoy spaghetti" may be equivalent to "Spaghetti pleases us". Furthermore, it is obvious that FLEX cannot be put to use until an automatic



routine is available for determining the antecedents of all pronouns, other than interrogative pronouns (here, e.g., interrogative "who" will be considered to be a perfect synonym for any personal name or description). Other projects have so far not come up with a solution for this problem, but it must be solved.

j. On semantic matching. The need for more than indexes of synonymy is illustrated by a sentence-pair like this: "This cobbler does poor work" and "The shoes made by this man are inferior." Here we need to show that "cobbler" somehow contains simultaneously elements which match both "shoes" and "made". Work has hardly begun on this point, and much more is needed.

### Bibliography

1. First Quarterly Progress Report on the Automation of General Semantics; Fred W. Householder jr., Principle Investigator; May 1, 1960.
2. Third Quarterly Report on the Automation of General Semantics; Fred W. Householder jr. and John Lyons, Principal Investigators; November 30, 1960.
3. Fourth Quarterly Report on the Automation of General Semantics; Fred W. Householder jr. and John Lyons, Principal Investigators; February 28, 1961.
4. Fifth Quarterly Report on Automatic Language Analysis; Prepared by John Lyons; May 31, 1961.
5. Sixth Quarterly Report on Automatic Language Analysis; Fred W. Householder jr. and James Peter Thorne, Principal Investigators; October 31, 1961.
6. Seventh Quarterly Report on Automatic Language Analysis; Fred W. Householder jr. and James Peter Thorne, Principal Investigators; March 1, 1962.

APPENDIX I -- Card Formats

<u>FORMAT #</u>	<u>FORMAT NAME</u>	<u>COLUMNS</u>	<u>FIELD NAME</u>
1	Dictionary	1-25 26-40 41-55	Alphabetic Item Dictionary Code Dictionary Order #
2	Text	1-71 72-80	Text Items Card Sequencing

Best Available Copy

# APPENDIX II -- File Descriptions

FILE #	NAME	CHARACTERISTICS	WORD	CHARACTER	FIELD NAME
1	Dictionary	Block = 500 Words	1	1-	Alphabetic Item (Zero Padding)
		Record = 20 Words	4	6	
		Header/Trailer	5	1-	Dictionary Code
2	Text	Block = 500 Words Record = 10 Words Header/Trailer	7	3	
			1	1-	Text Item (Zero Padding)
			4	6	
			5	1-2	Identification of Text
			5	6	Capitalization
			6	1-	Text Item Identification
			7	4	
3	Appended Text	Block = 500 Words Record = 20 Words Header/Trailer	1	1-	Text Item
			4	6	
			5	1-2	Text Identification
			5	6	Capital Letter Code
			6	1-	Item Identification
			7	4	
			7	5-6	Affix Code
			9	1-	Dictionary Match
			11	6	
			12	1	Noun/Pronoun 1=Noun 2=Pronoun 0=neither
				2	Verb/Conjunction 1=verb 2=Conjunction 0=neither

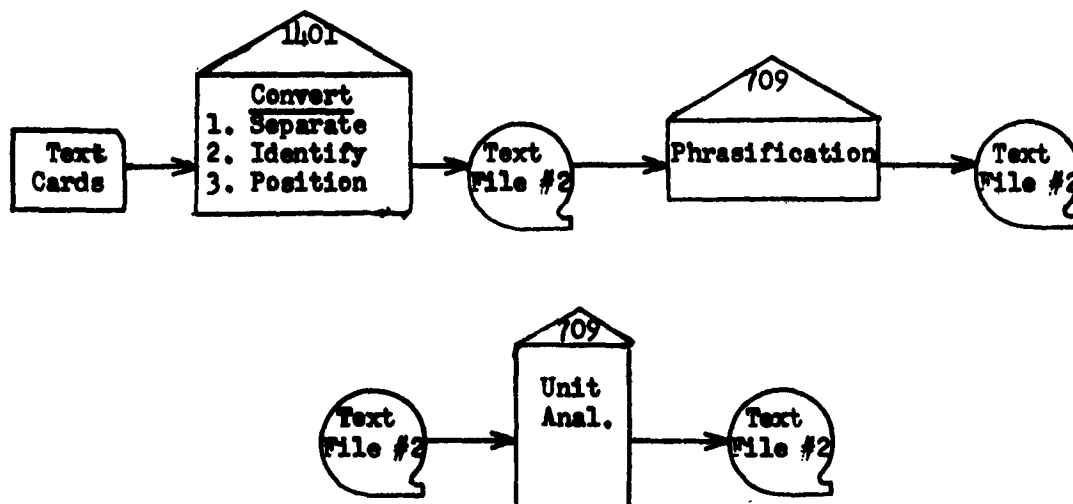
# APPENDIX II (Cont'd)

FILE #	NAME	CHARACTERISTICS	WORD	CHARACTER	FIELD NAME
				3	Adjective/Preposition 1=Adjective 2=Preposition 3=neither
				4	Adverb Code 1=Adverb 0=not an Adverb
			12	5-6	Noun/Pronoun Subclass
			13	1-2	Verb/Conjunction Subclass
			13	3	Adj./Preposition Subclass
			13	4	Adverb Subclass
			20	1-5	Dictionary Code
4	Error	Block = 20 Words Record = 10 Words Header/Trailer	= Record Format 1, File #2		
5	Resolved Text (1)	Block = 500 Words Record = 20 Words	1	1-}	Same as Record Format 1, File 3
			14	6 }	
			15	1-}	Indicator Codes
			17	6 }	
			18	1-6	Resolved Part of Speech
			19	1-3	Resolving Rule #
			20	1-5	Dictionary Code
			20	6	Ambiguity Code 1=N/Present Tense 2= -ing 4=Adj./Past Tense 8=Adj./Present Tense 16=N/Past Tense 32=Idiosyncratic Distribution

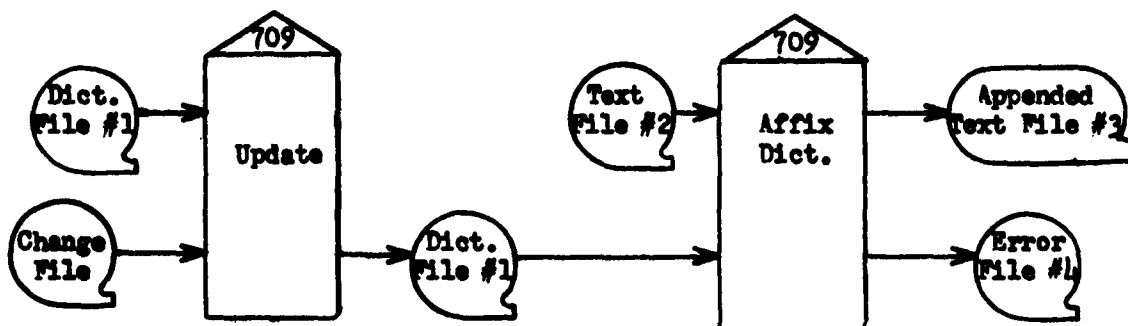
III

APPENDIX -- File Development

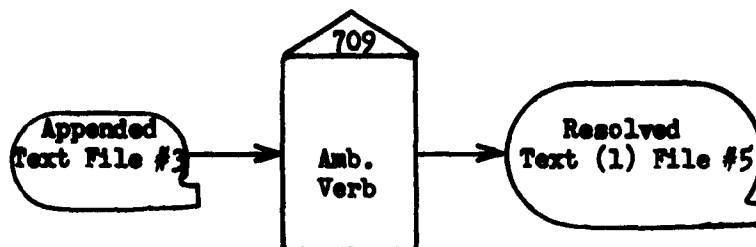
Block 1



Block 2



Block 3



#### APPENDIX IV --- Analyzable Unit Schema

##### Unit Analysis Rules

Note: I = item under consideration.  
P = item preceding I.  
#I and #P = unit numbers of I and P.

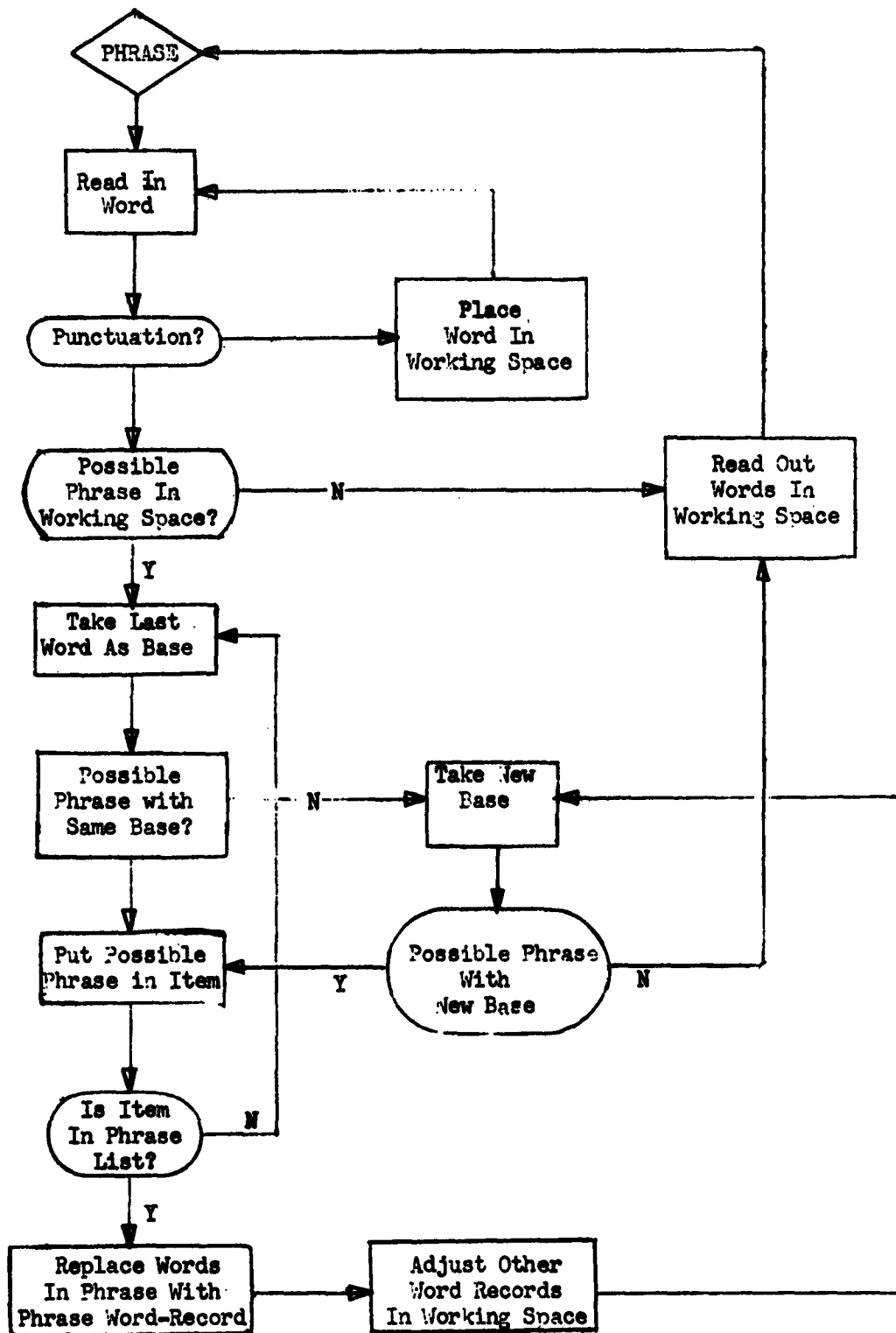
Two dashes are said to be paired if there are no colons, semi-colons between them and no parenthesis (or an even number of parenthesis) between them.

0. In the rules which follow, if I is the 1st item in the sentence assume #P = 0.
1. If I is a left parenthesis,  $\#I = \#P + 1$ .
2. If P is a right parenthesis,  $\#I = \#P - 1$ .
3. If I is the 2nd of paired dashes,  $\#I = \#P + 1$ .
4. If P is the 1st of paired dashes,  $\#I = \#P - 1$ .
5. Otherwise, take  $\#I = \#P$ .

##### Pictures

1.  $I = ( \Rightarrow \#I = \#P + 1$
2.  $P = ) \Rightarrow \#I = \#P - 1$
3.  $I = \text{2nd of paired dashes} \Rightarrow \#I = \#P + 1$
4.  $P = \text{1st of paired dashes} \Rightarrow \#I = \#P - 1$
5.  $\#I = \#P$ .

APPENDIX -- Phrasification Flow





# APPENDIX VI -- Form Subclass Codes

## Noun Subclass Codes

If an item belongs to the noun form class, then 03, 10, 17, 22, 23, and 25 indicate that the item is a non-countable item. The following table shows what noun subclasses are indicated by 01, 03-20, and 22-25. An "X" in a column opposite a number indicates that the subclass indicated by that number has the property named at the top of the column. The lack of an "X" indicates that the subclass does not have that property.

	COUNTABLE	UNCOUNTABLE	PROPER	MASCULINE	FEMININE	NEUTER
01	X					X
03		X				X
04	X		X			X
05	X			X		
06	X				X	
07	X			X	X	
08	X		X	X		
09	X		X		X	
10		X	X	X	X	
11	X			X	X	X
12	X	X		X		X
13	X		X	X		X
14	X	X	X	X	X	X
15	X		X	X		X
16			X		X	X
17		X	X	X		
18	X	X		X		
19	M O N T H S					
20	X				X	X
22		X	X	X	X	X
23		X		X		
24	d a t e s					
25		X	X	X		X

APPENDIX VI -- (Cont'd)

Verb Subclass Codes

The following table is analogous to the one used to explain noun subclasses:

	trans- itive	B/V = transitive=NP verb = two objects	transitive +to+verb	intrans- itive	copul- ative	transitive +adjective
01	X					
02	X			X		
03				X		
04	X		X	X		
05	X		X	X		X
06	X			X		X
07	X	X		X		
08	X	X	X	X		X
09	X		X	X		
10	X	X				
12	X	X	X			
13	X			X	X	
14	X	X	X	X		
15	auxiliaries					
16		do.	will.		can	
17	X		X		X	X
18	X		X			
19		X		X	X	
20	X			X		
21	X			X		
22	X		X	X		
24					X	
25	X	X		X		X
26	X			X		
27	X			X		X
28		X		X		
29	X	X		X		
30	X	X	X	X		
31	X	X	X			
32		X		X		

## APPENDIX VI --(Cont'd)

### Pronoun Subclass Codes

If an item belongs to the pronoun form class, then:

- 01 = object pronoun
- 02 = subject pronoun
- 03 = object/subject pronoun ambiguous item
- 04 = indefinite pronoun
- 05 = possessive pronoun
- 06 = reflexive pronoun
- 07 = English numeral

### Adjective Subclass Codes

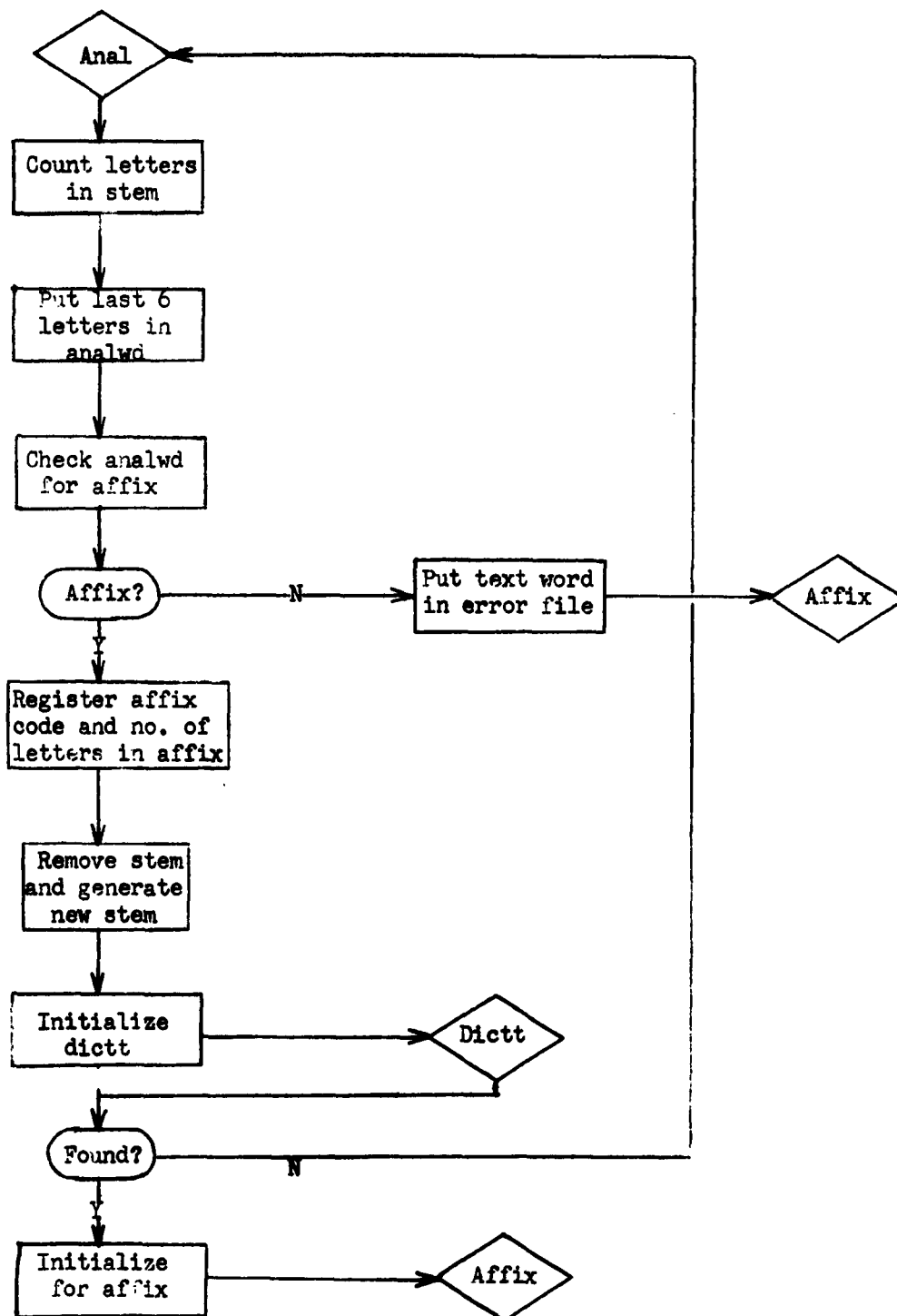
If an item belongs to the adjective form class, then:

- 01 = indefinite article
- 02 = Arabic numeral
- 04 = definite article (= the, no)
- 08 = possessive adjective (including his)

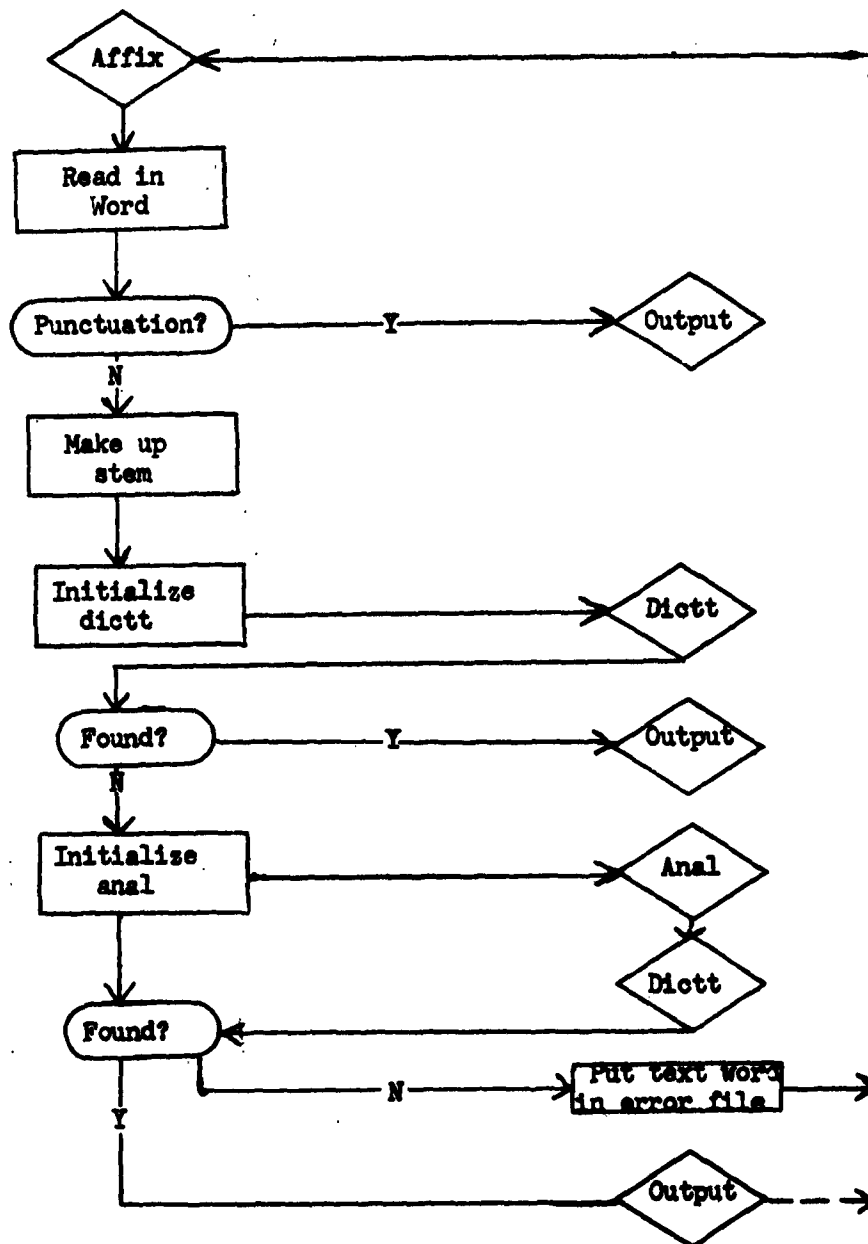
APPENDIX VII -- Dictionary Code

CHARACTER	DESCRIPTION	PARAMETER
1	Affix of item indicated	1
2	Phrase classification according to certainty	0 = always a phrase 1 2 3 4 } = type of uncertainty
3	Irregular forms	0 = not irregular form 1 = past tense other than 7 or 8 2 = past/present ambiguous 3 = plural 4 = singular/plural 5 = comparative 6 = superlative 7 = past participle only 8 = past tense only
4	Priority code	1 = noun 2 = verb 4 = adjective 0 = no priority
5	Absolute breaker code	1 = class B, breaker 3 = class A, breaker

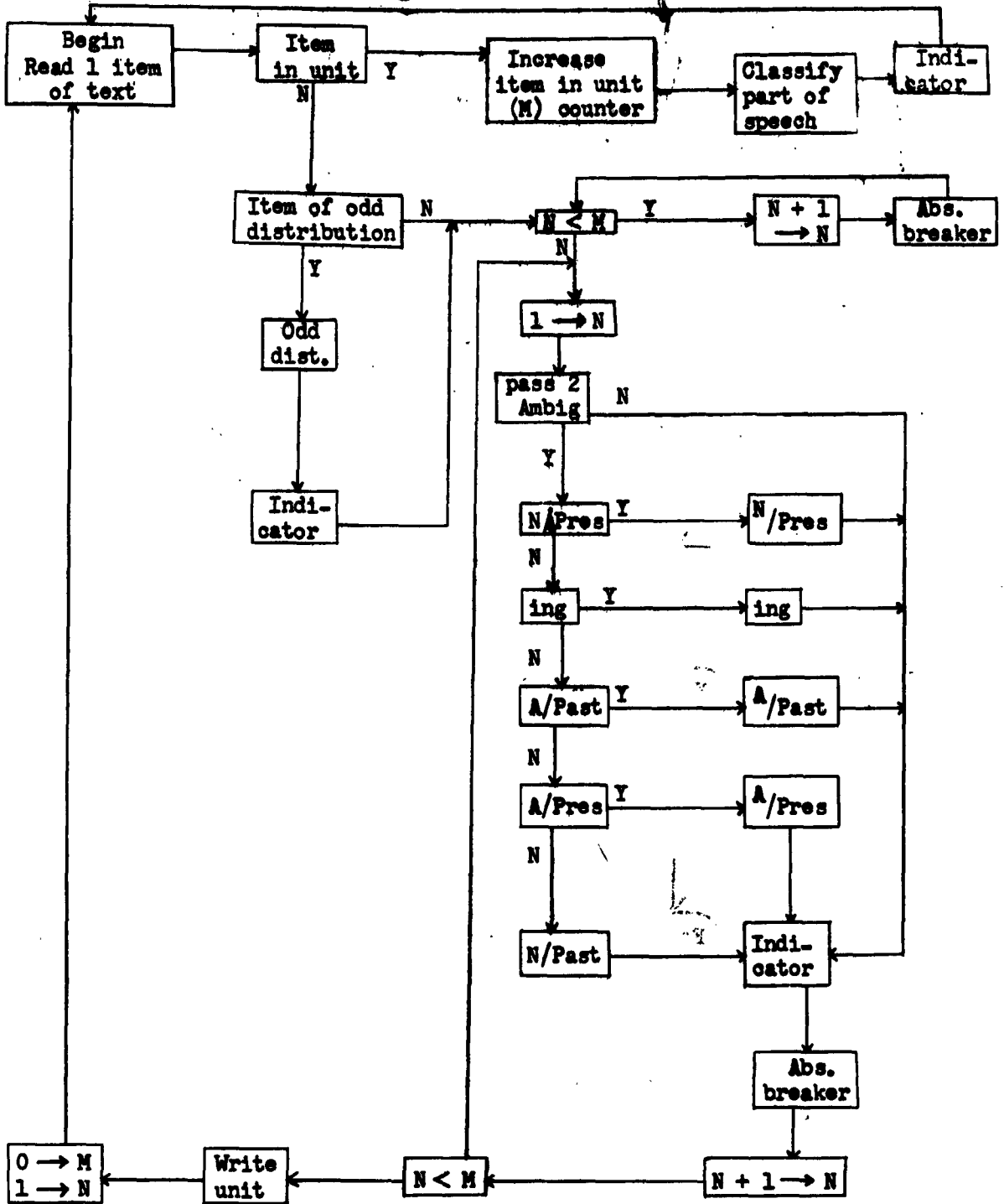
VIII  
APPENDIX -- Affix Dictionary Flow



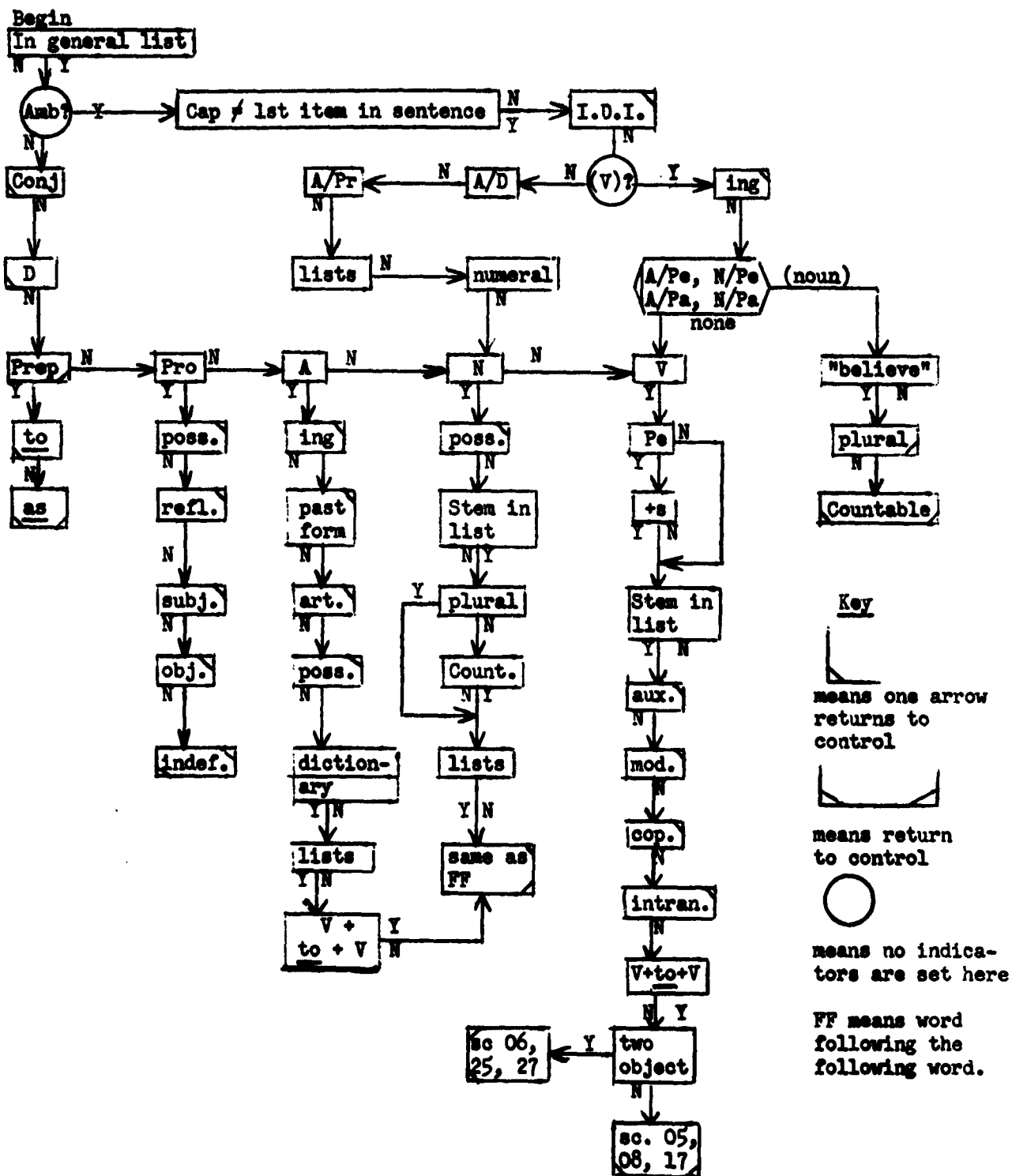
APPENDIX <sup>VII</sup> -- (Cont'd)



**APPENDIX I -- Control Flow Chart**



# APPENDIX -- Indicator Flow Chart





APPENDIX XI -- The Indicator Codes

0. V, V/A. 1.21
1. plural (N). 1.33
2. Cap. ≠ beginning of sent., subj. Pr., 1.1, 2.1, 8.1, +1, -1, Z
3. V subclass 05, 08, 17, 4.16
4. singular countable (N). 1.42, 1.44, 2.26, 8.10, 8.29
5. V, subject Pr. Z
6. present tense + Ø. A, AC, Z
7. thought, spoke. +14
8. Past (V) or past V and past participial adjective. +25, -5, Z
9. auxiliary, modal, copulative V. 1.12, 1.24, 2.6, 2.18, 4.1, 4.7, 8.10, 8.24, 8.26, +12
10. "Believe" Verbs. 1.30, 1.40
11. (N), -ing. 1.16
12. Intransitive V. 1.15
13. A, V = V subclass 06, 25, 27, 8.13
14. present V, auxiliary, modal, copulative V. 2.7
15. part of appear, consider. 2.19
16. two object V 1.23, 1.28, 4.16, 8.13, +21
17. V part of begin, continue, end, finish, open, stand, start, stop. 2.33
18. V part of begin, continue, help, keep, lie, send, sit, stand, start, stop, try; worth, 2.40
19. consisting, dying, speaking, talking, telling, thinking 2.21
20. V part of ascribe, attach, attribute, belong, cling, commit, convert, oppose, pertain, reconvert, relate, subject 1.9, -12

21. (as nominal is) according, alien, amenable, antagonistic, attributable, basic, complementary, contradictory, convert(s), equivalent, FF, foreign, hostile, inimical, liability, opposite, proportion(al), regard(s), resistant, respect(s), sensitive, similar, supplementary. 1.9, -12
22. exercise(s), method(s), procedure(s), technique(s) as N's 2.9
23. accident, aim, art, capable, certainty, custom, device, difficulty, ease, effect, feasibility, habit, hope, hopelessness, idea, impossibility, incapable, interest, means, method, necessity, object, obligation, possibility, practice, problem, purpose, question, result, rule, sake, and their plurals. 2.30
24. cure, device, facility, fame, flair, machine, necessity, need, notoriety, notorious, reason, talent and their plurals. 2.32
25. attributable, attribute(s)(d), belong, opposed, opposite, proportion(al), similar, subject(ed). 8.27
26. V ≠ (ascribe, attach, attribute, belong, cling, commit, convert, oppose, pertain, reconvert, relate, subject. -6)
27. modal, I, we, you, they, who. -6
28. A, V = subclasses of trans. V + to + V.
29. (N), A, Pr. 2.34
30. singular N. 1.36
31. plural N. 1.21, 1.23, 2.12, 8.12, +22, Z
32. singular countable N. 1.19, 1.20, 1.29, 1.41, 2.22, 4.15, +18, +20
33. N, (A) 1.15
34. article, Pr/A, possessive N or A 2.2
35. A, N, subject or indefinite Pr. who, whoever, no one, it, you 4.13

APPENDIX XI (con't)

0. and, nor, or
1. N, object Pr. you, it. 4.16
2. possessive A, whose. 2.20
3. present V +s, P. -3
4. V, P. 1.18, 8.20, 8.21, +16
5. V, P, article. 8.15
6. V, P, poss. A. 4.21
7. V, P, C. +20
8. V, N. 1.43
9. ing or past form adjective. 1.11, +20
10. D, A/D, A ≠ (article, possessive A). XLOOK4
11. poss. N, A, Pr/A. 1.28
12. A ≠ ing. 1.3, 2.5, 4.4, +10
13. V, N, article, Pr. Z
14. A, poss. Pr., his. -13
15. N, A. 1.13, 1.20, 2.29, +25, Z
16. D, Pr/A. 4.24
17. D, A/D XLOOK2, Z
18. D, A/D, Pr/A (≠ another, each, no one, one, this; few, many, several, these, those or numeral). XLOOK3
19. down, except, like, near, till. -9
20. poss. N or A, article. 1.2, 2.28, 4.4, 8.3, +2, -6
21. V, sub. Pr., it. Z
22. V, A. 1.9, +27
23. poss. N or A or Pr, whose, 4.5, +17
24. V ≠ (modal, auxiliary), N, Pr/A, P, obj. or indef. Pr., Z

25. article, poss. A, Pr/A, indefinite Pr., whoever, no one. 8.13
26. down, near, still, well. -5
27. Pr/A. 2.37, 4.21
28. art., A/Pr., poss. Pr., his, whose, that, which 1.29
29. another, each, no one, one, this 1.18, +15
30. Pr/A, that +15, +16
31. Pr/A, more, most. 8.15
32. another, each, no one, one, this; few, many, several, these, those,  
1.18
33. another, any, each, much, no one, one, other, such, this. 8.8
34. poss. or refl. or obj. Pr., it, you. 8.7
35. article, A/Pr, indef. or refl. Pr, more, most, very. +24

APPENDIX XI (con't)

0. with, between 2.31
1. that, which. 4.24
2. more, most. 8.31
3. most, too. 2.25
4. comparatively, strikingly, too, very. 8.4
5. still, well. -7
6. art., subj. or indef. or poss. Pr, poss. A, plural N, proper N, verb,  
more, most, that Z
7. article. 1.29, 2.3, 2.22, 4.2, 4.3, -3, -5
8. like, except. -6
9. since, until Z
10. P ≠ to. 1.8, 2.24, 8.12, +9, Z
11. P ≠ as. 4.9
12. (N), A. 4.20, 4.25
13. symbol, Arabic numeral. 1.24
14. symbol, Arabic or written numeral. 4.23
15. subj. or indef. or poss. Pr, who, whoever, it, you. 1.5
16. Arabic or written numeral. 1.29, 1.30, 1.33, +23
17. object or indefinite or poss. Pr, whoever, no one, it, you 1.6
18. object or indef. Pr, who, no one, it, you. 8.19
19. object or indef. Pr, more, most, it, you. 2.11
20. object or indef. Pr, no one, it, you. 1.27, 4.10, +3
21. N, object or indefinite Pr, you, it. +21
22. subj. or indef. Pr, who, whoever, no one, it, you. 4.8
23. subject or indefinite Pr, who, whoever, it, you. +5
24. N, A/Pr, indef. or subj. Pr, it, you.

25. plural N or A/Pr or subj Pr, you. AC
26. indef. or poss. Pr, poss A, present A, dictionary A, such, sense, reason. Z
27. indefinite Pr, who, whoever, no one. 2.13, 2.14
28. A, singular countable N. Z
29. indef. Pr, who, you. 8.20
30. object or indefinite Pr., it, you. 2.15
31. although, because, since, though, when; how, if, until, where, whereas, whether, whilst; lest, till, unless, whenever, wherever, why. 1.25
32. although, because, since, though, when; as, how, however, once, so, that, thus, what, whatever, wherever. 4.19
33. after, before, besides, since, until. Z
34. although, because, since, though, when; how, if, until, where, whereas, whether, whilst, while; what. 4.22, 8.5
35. after, before, since, until. Z

## APPENDIX XII

### The Search Routines

The ambiguity resolution routines require information about preceding items and about following items. Before such an item can be interrogated the programmer must first determine if there exists an item to be interrogated. Usually the programmer considers that there is no such item in the unit if punctuation is encountered. Therefore all of the routines exit NO WORD FOUND whenever punctuation intervenes between the item with which the search begins search item and the item looked for. For example, suppose we want to know if the item in front of the search item is to. If the item in front of the search item is a comma the exit is NO WORD FOUND, no matter how many words in unit there are preceding the item. However, the address of the preceding (or following) item is always given, whether it is a punctuation mark or not. (This is to handle certain special cases.) If it is punctuation, the address word has a negative sign.

The calling sequence for each search subroutine has the following format,

	TSX	<del>XL</del> XY,4
A	PZE	ATWR
B	PZE	NWF
C	PZE	**

Here X = P for searching left for a preceding item; X = F for searching right for a following item.

Y = 1 means nothing is ignored,

Y = 2 " adverbs and adjective/adverbs are ignored,

Y = 3 " adverbs and adjective/adverbs and pronoun-adjectives  
# (this, etc.) are ignored,

Y = 4 " adverbs and adjective/adverbs, and adjectives # (article,  
possessive).

For example: FL00K2 looks for the first following word which is not an adverb or an adverb/adjective.

Transfer is made indirectly to the address (say) "NWF" in "B" if NO WORD FOUND.

If WORD FOUND (i.e. a word which the programmer will want to test) transfer is made directly to "C" + 1.

"C" is always filled in by the search routine. Minus zero is stored there if there are only ignored items between the search item and the beginning or end of the unit, depending whether a PL00K or a FL00K is used. If a punctuation mark is found, exit is made to "NWF" and its address is placed in "C".

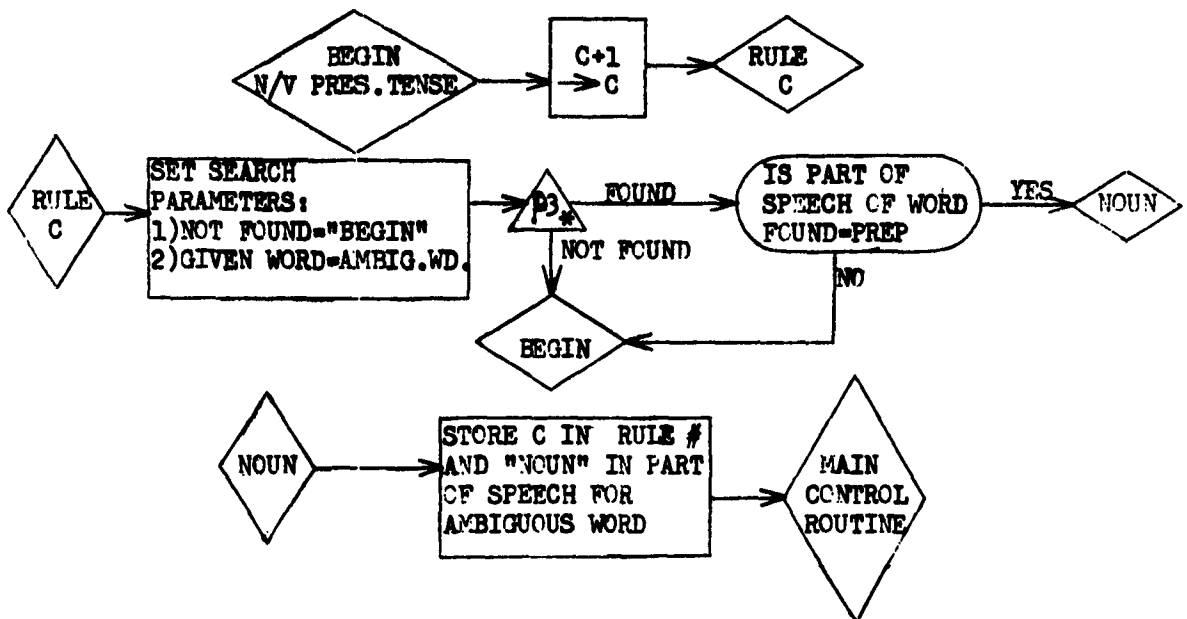
"A" must be filled in by the programmer. The programmer must store in "A" the first address of the text item record. "A" will always be assembled as PZE\*\*.



APPENDIX XIII--Typical Ambiguity Routine Rule (noun/verb present tense)

Rule C

Is the ambiguous word preceded (ignoring adv/prep and adj/pronouns) by a preposition? If yes assign as a noun. If no, or if no word precedes, apply next sequential rule.



\*P3 = PL00K3 (see Appendix XII - The Search Routines)

## APPENDIX XIV -- The Ambiguity Routines

### Present Tense Verb-Noun Resolution

Note W = word, P = previous word, F = following word, (P) = previous word, ignoring adverbs and adverb-adjectives, /P/ = same as (P) but also ignoring all adjective-pronouns except this, each, another, one, these, those, several, many, few, no one, and English numerals.

1. Is W capitalized and not the 1st word in the sentence? If yes, take W as NOUN.
2. Is /P/ an article, possessive N, or possessive adjective? If yes, take W as NOUN.
3. Is (P) an adjective ing? If yes, take W as NOUN.
4. Is (P) a modal verb? If yes, does W have s as affix (final)? If yes, take as NOUN; if no, take as VERB.
5. Is (P) a subject, indefinite, or possessive pronoun, you, it, who, or whoever? If yes, take as VERB.
6. Is (F) an object, indefinite, or possessive pronoun, you, it, no one, whoever? If yes, take as VERB.
7. Is (F) = of? If yes, take as NOUN.
8. Is (P) a preposition to? If yes, take as NOUN.
9. Is P = to? If yes, is W a+s (i.e. has it the affix s)? If yes, take as NOUN; if no, see if PP = (as nominals) according, alien, amenable, antagonistic, attributable, basic, complementary, contradictory, contrary, convert, converts, equivalent, W, foreign, hostile, imideal, liability, opposite, proportion, proportional, regard, regards, resistent, respect, respects, sensitive, similar, or supplementary or if previous verb or adjective is part of ascribes, attach, attribute, belong, attain, cling, commit, convert, oppose, pertain, reconvert, relate, or subject. If yes, take as NOUN. If no, take as VERB. If no PP, take as VERB.
10. Is (P) = to? If yes, is W a+s. If yes, take as NOUN.
11. Is P = that? If yes, is (P)P a verb? If yes, take W as NOUN; if no, is (P)P an -ing or past form adjective? If yes, is (P) (P)P an auxiliary? If yes, take as NOUN. If no, is PP a preposition? If yes, is W a+s? If yes, take as VERB; if no, take as NOUN.
12. Is (F) an auxiliary, modal, or copulative verb? If yes, take as NOUN.
13. Is (F) a past tense/adjective? If yes, is (F)(F) something which may be a noun or an adjective? If no, take as NOUN.

14. Is (F) a verb? If yes, is W intransitive as verb. If yes, take as NOUN.
15. Is (P) an intransitive verb? If yes, take as NOUN.
16. Is F a candidate noun or an -ing? If yes, is (F)F a verb? If yes, take as NOUN.
17. Is (P) = to? If yes take as NOUN.
18. Is (P) = this, each, another, one, no one, these, those, several, many, few? If yes, is P(P) a preposition or verb? If yes, take as NOUN; if no, is W +s? If yes, is (P) = this, each, another, one, no one? If yes, take as VERB; if no, take as NOUN. If W is not +s, is (P) = this, each, another, one? If yes, take as NOUN. If no, take as VERB.
19. Is P a singular countable noun? If yes, are there (if anything) only adverbs, adverb/adjectives, adjectives (≠ article, possessive adjective), and other singular countable nouns between P and the next Previous punctuation? If yes, take as NOUN.
20. Is W +s? If no, is F a singular countable noun? If yes, is FF a noun or an adjective? If no, take as NOUN.
21. Is F a plural noun? If yes, is (F)F a verb or verb/adjective? If yes, take ambiguous item as NOUN.
22. Is P a noun? If yes, is (ignoring Adjectives also) (P)P a verb? If yes, take W as NOUN.
23. Is (P) a plural noun? If yes, is Previous Verb a two-object verb? If yes, take W as NOUN; if no, take as VERB; if no previous verb, take W as VERB.
24. Is (W) first word in unit? If yes, has W +s? If yes, take W as NOUN; if no, see if F is an Arabic numeral or other symbol? If yes, take W as NOUN. If no, is there a Following Verb? If no, take W as VERB; if yes, see if verb is auxiliary, modal, or copulative. If yes, take W as NOUN.
25. Is P = although, though, because, how, if, lest, unless, since, till, until, when, whenever, where, wherever, whether, while, whilst, whereas, or why? If yes, take W as NOUN.
26. Is (F) an adjective? If yes, is F(F) a preposition? If yes, take as NOUN.
27. Is (P) an object or indefinite pronoun or no one? If yes, is (P)(P) = let? If yes, take as VERB.
28. Is (F) an adjective, adjective/pronoun, or possessive noun? If yes, is there a verb between W and the Previous punctuation? If yes, is it a two-object verb? If no, take as VERB. If there is no verb, take W as VERB.

29. Is (P) an English numeral? If yes, is W+s? If yes, is P(P) a singular countable noun? If yes, is PP(P) an article? If no, take as VERB; if yes, or if P(P) is not a singular countable noun, take as NOUN.
30. Is (F) = that? If yes, is (F)(F) a candidate verb? If yes is W a verb which takes a that-clause as object (i.e. a "believe" verb)/ If yes, take as VERB.
31. If (F) an English numeral? If yes, take W as VERB.
32. Is W capitalized? If yes, take as NOUN.
33. Is (P) an English numeral? If yes, is F a noun candidate? If yes, is F a plural noun candidate? If yes, take as NOUN; if no, see if FF is a plural noun candidate? If yes take as NOUN; if no, see if P(P) is a singular countable noun. If no, take as VERB; if yes, take as NOUN.
34. Is W+s? If no, is (P) an adjective? If yes, take as NOUN.
35. Is W a+y? If yes, is (P) a noun? If yes, is P(P) an indefinite article? If yes, take as VERB.
36. Is W a+s? If yes, is P a singular noun? If yes, is F a singular noun? If yes, take as VERB.
37. Is W first word in unit? If yes, take as NOUN.
38. Is P an adjective? If yes, take as NOUN.
39. Is P = that? If yes, take as VERB.
40. Is (F) = that? If yes, is F(F) a preposition. If yes, take as VERB; if no, see if W is subclass 07, 08 10-12, 14, 19, 25, 28-32 (i.e. "believe" verbs). If no, take as NOUN. If yes, take as VERB.
41. Ignoring everything but punctuation and verbs find Previous preposition. If found, see if this is only followed by singular countable nouns or adjectives ≠ (article, possessive adjective). If yes, take as NOUN.
42. Is (P) a conjunction? If no, go to AND CONCORD. If yes, is F a preposition? If yes, might W be a singular countable noun? If yes, take as VERB; if no, see priority. If no priority, go to AND CONCORD.
43. Is P a comma? If yes, is F a comma? If yes, is PP a noun or verb? If yes, take same as PP.
44. Is P a comma? If yes, is F a preposition? If yes, is W a singular countable noun candidate? If yes, take as VERB; if no, see priority.

## CONCORD

- I. If there is a priority, take accordingly as NOUN or VERB.
- II. Find Preceding verb, absolute breaker, or punctuation, or beginning of unit, call this (P.B.). Ignoring everything else, find first subjects\* or preposition following this (P.B.), call this S. If S<sub>1</sub> is a preposition, take W as NOUN. If S is a subject, go to A. If neither is found, go to III.
  - A. Is (P.B.) a verb? If yes, is W a-s? If ~~yes~~, is (F) a plural noun candidate? If yes, take as VERB, if no is (F) a candidate verb? If yes, take as NOUN, if no, go to part A of AND CONCORD.
    1. Is (P.B.) of subclass 08, 09, 12, 19-22, 26-30 (trans. + to + verb)? If yes take as VERB. If no is (P.B.) an auxiliary? If yes is (F) (P.B.) an adjective with the above verb subclasses? If yes, take as VERB. If no, go the AND CONCORD.
  - B. Does and, or, or nor occur between (P.B.) and W? If no, is S a plural? If yes, is W plural as noun? If yes, take as NOUN; if no, take as VERB. If S is not a plural, see if W is plural as noun. If yes, take W as VERB; if no, as NOUN.
- III. Take as NOUN.

\* subject = noun, subject pronouns (2+3), indefinite pronouns and the adjective/pronoun.

## AND CONCORD

Is there a Preceding Verb, P. V., in the unit? If no, take as NOUN, if yes, is P. V. a present tense + s? If yes, is W a + s? If yes, take as NOUN; if no, take as VERB.

A. Find the first subject, S, Preceding P.V.

1. If no subject found, take W as NOUN
2. If subject found, look for preposition, P, Preceding S (ignoring everything but prepositions, verbs, punctuation, and absolute breakers). If no P, do concord\* with this subject.
3. If P found, look for first subject preceding it and go back to step 1.

### \*concord

plural subject + . . . + W a + <u>s</u>	W = NOUN
non plural subject + . . . + W a + <u>s</u>	W = VERB
plural subject + . . . + W not a + <u>s</u>	W = VERB
non plural subject + . . . + W not a + <u>s</u>	W = NOUN

# Present Tense/Noun Ambiguities Pictures

Note: "+" = immediately next word; "(+)" = next word disregarding adverbs and adverb-adjectives; "/+/" = same as (+) but also ignoring all adjective-pronouns except this, each, another, one, etc.; "W" = ambiguous item; ( ) around anything except + means "candidate for"; "N" = NOUN; "V" = VERB; "A" = Adjective; "D" = ADVERB; etc; "\$" = unit boundary.

1. W = capitalized ≠ 1st word in sentence  $\Rightarrow$  W = NOUN.
2. art. }  
poss. N }  $/+/W \Rightarrow$  W = NOUN  
poss. A }
3. A ≠ ing (+) W  $\Rightarrow$  W = NOUN
4. Modal V (+) W = a + s  $\Rightarrow$  W = NOUN  
Modal V (+) W ≠ a + s  $\Rightarrow$  W = NOUN
5. subj. }  
indef. } Pron. you, it, who, whoever (+) W  $\Rightarrow$  W = VERB  
poss. }
6. W (+) obj. }  
indef. } Pr., you, it, whoever, no one  $\Rightarrow$  W = VERB  
poss. }
7. W (+) of  $\Rightarrow$  W = NOUN
8. Prep ≠ to (+) W  $\Rightarrow$  W = NOUN
9. to + W = a + s  $\Rightarrow$  W = NOUN  
according, etc. + to + W ≠ a + s  $\Rightarrow$  W = NOUN  
V = ascribe, etc. + ... + to + W ≠ + s  $\Rightarrow$  W = NOUN  
A  
to + W ≠ a + s  $\Rightarrow$  W = VERB
10. to (+) W = a + s  $\Rightarrow$  W = NOUN
11. V (+) that + W  $\Rightarrow$  W = NOUN  
aux. V. (+) A = "ed" or "ing" (+) that + W  $\Rightarrow$  W = NOUN  
Prep. + that + W = a + s  $\Rightarrow$  W = VERB  
Prep. + that + W ≠ a + s  $\Rightarrow$  W = NOUN
12. aux. }  
W(+) modal } V  $\Rightarrow$  W = NOUN  
copul. }
13. W (+) "ed" (+) NOUN or ADJ.  $\Rightarrow$  W = NOUN
14. W = intran. (V) (+) VERB  $\Rightarrow$  W = NOUN

15.  $V = \text{intran. } (+) W \Rightarrow W = \text{NOUN}$
16.  $W + \left\{ \begin{array}{l} N \\ N/V \\ \text{"ing"} \end{array} \right\} (+) V \Rightarrow W = \text{NOUN}$
17. to (+)  $W \Rightarrow W = \text{NOUN}$
18.  $\left. \begin{array}{l} \text{Prep} \\ V \end{array} \right\} + \text{this, etc. } (+) W \Rightarrow W = \text{NOUN}$   
 $\text{this, etc. } (+) W = a+s \Rightarrow W = \text{VERB}$   
 $\text{these, etc. } (+) W = a+s \Rightarrow W = \text{NOUN}$   
 $\text{this, etc. } (+) W \neq a+s \Rightarrow W = \text{NOUN}$   
 $\text{these, etc. } (+) W \neq a+s \Rightarrow W = \text{VERB}$
19. punctuation +  $\left\{ \begin{array}{l} \emptyset \\ \text{ADV/ADJ} \\ A \neq \text{art., poss. A.} \end{array} \right\} \dots \text{sing. count N} + W \Rightarrow W = \text{NOUN}$
20.  $W = a+s + \text{sing. count N} + \left\{ \begin{array}{l} N \\ V \end{array} \right\} \Rightarrow W = \text{NOUN}$
21.  $W + \text{plur. N } (+) \left\{ \begin{array}{l} V \\ A \end{array} \right\} \Rightarrow W = \text{NOUN}$
22.  $V (\text{ignoring A's}) (+) N + W \Rightarrow W = \text{NOUN}$
23.  $V = \text{two obj. } \dots \text{ plur. N } (+) W \Rightarrow W = \text{NOUN}$   
 $V \neq \text{two obj. } \dots \text{ plur. N } (+) W \Rightarrow W = \text{VERB}$   
 $\neq \text{VERB } \dots \text{ plur. N } (+) W \Rightarrow W = \text{VERB}$
24.  $\neq (+) W = a+s \Rightarrow W = \text{NOUN}$   
 $\neq (+) W \neq a+s \left\{ \begin{array}{l} \text{NOUN} \\ \text{SUBJ} \end{array} \right\} \Rightarrow W = \text{NOUN}$   
 $\neq (+) W \neq a+s \dots \text{no V occurs } \dots \neq \Rightarrow W = \text{VERB}$   
 $\neq (+) W \neq a+s \dots \left\{ \begin{array}{l} \text{aux.} \\ \text{modal} \\ \text{copul.} \end{array} \right\} V \Rightarrow W = \text{NOUN}$
25. although, et. +  $W \Rightarrow W = \text{NOUN}$
26.  $W (+) A + \text{Prep} \Rightarrow W = \text{NOUN}$
27.  $\left. \begin{array}{l} \text{obj.} \\ \text{let } (+) \text{ indef.} \end{array} \right\} \text{Pron, no one } (+) W \Rightarrow W = \text{VERB}$
28. punctuation  $\dots \left\{ \begin{array}{l} V \neq \text{two obj.} \\ \text{no verb occurs} \end{array} \right\} \dots W (+) \left\{ \begin{array}{l} A \\ A/\text{Pron} \\ \text{poss. N} \end{array} \right\} \Rightarrow W = \text{VERB}$



29.  $\neq$  art. + sing. count. N ;  $\left\{ \begin{array}{l} \text{NNN} \\ \text{Numeral} \end{array} \right\} (+) W = a+s \Rightarrow W = \text{VERB}$   
 $\neq$  art. + sing. count. N + 1 (+)  $W = a+s \Rightarrow W = \text{VERB}$   
 art. + sing. count. N. + Numeral (+)  $W = a+s \Rightarrow W = \text{NOUN}$   
 Numeral (+)  $W = a+s \Rightarrow W = \text{NOUN}$
30.  $W = a$  "believe" (+) that (+) (V)  $\Rightarrow W = \text{VERB}$
31.  $W (+)$  Numeral  $\Rightarrow W = \text{VERB}$
32.  $W =$  capitalized  $\Rightarrow W = \text{NOUN}$
33. Numeral (+)  $W$  + plural (N)  $\Rightarrow W = \text{NOUN}$   
 Numeral (+)  $W$  + (N) + plural (N)  $\Rightarrow W = \text{NOUN}$   
 $\neq$  sing. count. N + Numeral (+)  $W \Rightarrow W = \text{VERB}$   
 sing. count. N + Numeral (+)  $W \Rightarrow W = \text{NOUN}$
34. A (+)  $W \neq a+s \Rightarrow W = \text{NOUN}$
35. indef. art. + N (+)  $W = a+s \Rightarrow W = \text{VERB}$
36. sing. N +  $W = a+s$  + sing. N  $\Rightarrow W = \text{VERB}$
37.  $\neq + W \Rightarrow W = \text{NOUN}$
38. A +  $W \Rightarrow W = \text{NOUN}$
39. that +  $W \Rightarrow W = \text{VERB}$
40.  $W (+)$  that + prep.  $\Rightarrow W = \text{VERB}$   
 $W \neq V$  subcl. "believe" (+) that +  $\neq$  prep.  $\Rightarrow W = \text{NOUN}$
41.  $\left. \begin{array}{l} V \\ \text{punctuation} \end{array} \right\} \dots \text{Prep} + \left\{ \begin{array}{l} A \neq \text{poss. A, art., etc.} \\ \text{sing. count. nouns} \end{array} \right\} + W \Rightarrow W = \text{NOUN}$
42. Conj. (+)  $W =$  sing. count. (N) + prep.  $\Rightarrow W = \text{VERB}$   
 conj. (+)  $W \neq$  sing. count. (N) + prep.  $\Rightarrow$  Go accordingly to priority  
 (No priority, go to AND CONCORD)
43. N + , + W + ,  $\Rightarrow W = \text{NOUN}$   
 V + , + W + ,  $\Rightarrow W = \text{VERB}$
44. ,  $W =$  sing. count. (N) + prep.  $\Rightarrow W = \text{VERB}$   
 ,  $W \neq$  sing. count. (N) + prep.  $\Rightarrow$  Go accordingly to priority  
 (No priority, go to next rule)

CONCORD -- to go with RIMEAA

(NOTE: For definition of "(P.B.)", "S" and "subject" see write up of CONCORD.)

- I. W = priority = 0    Go to II, Priority = 1    W = NOUN, Priority = 2    W = VERB.
- II. (P.B.) . . . S = prep  $\Rightarrow$  W = NOUN  
(P.B.) . . . S = subject  $\Rightarrow$  Go to A.  
(P.B.) . . . no S found . . . W  $\Rightarrow$  Go to III.
- A. (P.B.) = V . . . W = +s (+) plural (N)  $\Rightarrow$  W = VERB.  
(P.B.) = V . . . W = +s (+) (V)  $\Rightarrow$  W = NOUN.  
(P.B.) = V . . . W = +s  $\Rightarrow$  go to A of AND CONCORD.
- B. (P.B.) . . . S = plural . . . (no and, or, nor) . . . W = plur(N)  $\Rightarrow$  W = NOUN  
(P.B.) . . . S = plural . . . (no and, or, nor) . . . W  $\neq$  plur(N)  $\Rightarrow$  W = VERB  
(P.B.) . . . S  $\neq$  plural . . . (no and, or, nor) . . . W = plur(N)  $\Rightarrow$  W = VERB  
(P.B.) . . . S  $\neq$  plural . . . (no and, or, nor) . . . W  $\neq$  plur(N)  $\Rightarrow$  W = NOUN
- III. W = NOUN

# . . . (no verb) . . . Conj. (+)  $W \Rightarrow W = \text{NOUN}$

# . . . (+) V . . . Conj. (+)  $W = a+s \Rightarrow W = \text{NOUN}$

# . . . (+) V . . . Conj. (+)  $W \neq a+s \Rightarrow W = \text{VERB}$

A.

1. # . . . (no subject) . . . V . . . C (+)  $W \Rightarrow W = \text{NOUN}$

2. (no prep. occurs) . . . subject . . .  $W \Rightarrow \text{Full Concord}^*$   
with this subject

3. Repeat

\* as in CONCORD IIB.

## APPENDIX XIV (cont.)

### -ing Ambiguities Rules

NOTE: F(F) means "word following (F)", where (F) means "word following W" (= ambiguous item), ignoring adverb/adjectives and adverbs.

P\* is like (P) except that it also ignores adjectives.

1. Is W capitalized but not the first word in the sentence? If yes, take as NOUN.
2. Is (F) an article, pronoun/adjective, possessive adjective or possessive noun? If yes, take W as VERB.
3. Is (P) an article? If yes, take as ADJ.
4. Is P = very? If yes, take as ADJ.
5. Is (P) = adjective (≠ ing)? If yes, take as ADJ.
6. Is (P) an auxiliary, modal, or copulative? If yes, take as ADJ.
7. Is (F) a present tense verb, modal, copulative or auxiliary? If yes, take as ADJ.
8. Is (P) = during? If yes, take as ADJ.
9. Is F = method(s), procedure(s), exercise(s), or technique(s)? If yes, take as ADJ.
10. Is W almost always a verb (i.e. a L.D.P.)? If yes, take as VERB.
11. Is F an object or indefinite pronoun, more, most, you, it? If yes, take as VERB.
12. Is (P) a plural noun? If yes, take as VERB.
13. Is P an indefinite pronoun, who, whoever, no one? If yes, take as VERB.
14. Is F an indefinite pronoun, no one, who, whoever? If yes, take as VERB.
15. Is (P) an indefinite or object pronoun, it, you? If yes, is F a candidate noun? If no, take as VERB.
16. Is F = itself? If yes, is F a verb? If yes, is (P) a noun? If no, take as VERB.
17. Is F a reflexive pronoun? If yes, take as VERB.
18. Is (W) the first word in the unit? If yes, is next verb candidate part of auxiliary, modal, or copulative? If yes, take as ADJ.

19. Is P = as? If yes, is F punctuation or as? If yes take as ADJ; if no is F an adverb? If yes, take as ADJ; if no, see if Previous Verb is part of appear, consider. If yes, take as ADJ.
20. Is (F) a possessive adjective or whose? If yes, take as VERB.
21. Is (F) = of? If yes, is W = consisting, dying, speaking, talking, telling, thinking. If yes, take as VERB; if no, take as ADJ.
22. Is P a singular countable noun? If yes, is P\*P an article? If no, take as NOUN.
23. Is F = due to? If yes, take as NOUN.
24. Is (F) = preposition (≠ to)? If yes, take as VERB.
25. Is P = most, too? If yes, take as ADJ.
26. Is F a singular countable noun candidate? If yes, take as ADJ.
27. Is P a verb? If yes, take as ADJ.
28. Is P a possessive noun or possessive adjective? If yes, take as ADJ.
29. Is W intransitive as a verb? If yes, is F a noun or adjective? If yes, take as adjective; if no, take as VERB.
30. Is (P) = of? If yes, is P(P) = accident, aim, art, capable, certainty, custom, device, difficulty, ease, effect, feasibility, habit, hope, hopelessness, idea, impossibility, incapable, interest, means, method, necessity, object, obligation, possibility, practice, problem, purpose, question, result, rule, sake or their plurals? If yes, take as VERB; if no, as ADJ.
31. Is (P) = with, between? If yes, take as ADJ.
32. Is (P) = for? If yes, is P(P) = cure, device, facility, fame, flair, machine, necessity, need, notoriety, notorious, reason, talent, or their plurals? If yes, take as VERB.
33. Is (P) = by? If yes, is (P)(P) part of begin, continue, end, end up, finish, finish up, open, start, start out, start up, stop? If yes, take as VERB.
34. Is (P) a preposition? If yes, is F a punctuation mark? If yes, take as ADJ.; if no, is F an adverb? If yes, take as VERB; if no, see if F is a pronoun, adjective, or candidate noun. If no, take as ADJ; if yes, see if FF is a preposition. If yes, take as VERB; if no, see if P(P) is a verb. If yes, take as ADJ.
35. Is F a candidate noun? If yes, is (F)F a verb? If yes, take as ADJ.
36. Is (P) = let? If yes, take as ADJ.

37. Is (P) a pronoun/adjective? If yes, take as ADJ.
38. Is (P) a conjunction (i.e. and, but, or, nor, than) or a punctuation mark? If yes, is F a noun? If yes, is P(P) an adjective? If yes, take as ADJ.
39. Is (P) a punctuation mark or a conjunction? If yes, is P(P) an -ing word? If yes, take same as P(P).
40. Is F a punctuation mark? If yes, is Previous Verb part of begin, continue, help, keep, lie, send, sit, stand, start, stop, try, worth? If yes, take as VERB; if no, take as ADJ.
41. Take W as VERB.

NOTE: 1) No search crosses a punctuation mark.  
2) -ing's resolved as VERB are also set as ABSOLUTE BREAKERS.

-ing Ambiguities Pictures

1. W = capitalized ≠ first word in sentence ⇒ W = NOUN
2. 
$$W (+) \left\{ \begin{array}{l} \text{art,} \\ \text{pron/A} \\ \text{Poss. A.} \\ \text{Poss. N.} \end{array} \right\} \Rightarrow W = \text{VERB}$$
3. Art. (+) W ⇒ W = ADJ.
4. Very + W ⇒ W = ADJ.
5. A ≠ ing (+) W ⇒ W = ADJ.
6. 
$$\left. \begin{array}{l} \text{Aux.} \\ \text{Modal} \\ \text{Copula.} \end{array} \right\} (+) W \Rightarrow W = \text{ADJ.}$$
7. 
$$W (+) \left\{ \begin{array}{l} \text{Aux.} \\ \text{Modal} \\ \text{Copula.} \\ \text{Present Tense V.} \end{array} \right\} \Rightarrow W = \text{ADJ.}$$
8. During (+) W ⇒ W = ADJ.
9. W + Method(s), etc. ⇒ W = ADJ.
10. W = "funny word" ⇒ W = VERB
11. 
$$W + \left\{ \begin{array}{l} \text{obj. Pr.,} \\ \text{indef. Pr.,} \end{array} \right\} \left\{ \begin{array}{l} \text{more, most, you, it} \end{array} \right\} \Rightarrow W = \text{VERB}$$
12. Plur. N (+) W ⇒ W = VERB
13. 
$$\left. \begin{array}{l} \text{indef. Pr.} \\ \text{who,} \\ \text{whoever,} \\ \text{no one} \end{array} \right\} + W \Rightarrow W = \text{VERB}$$
14. 
$$W + \left\{ \begin{array}{l} \text{indef. Pr.} \\ \text{who,} \\ \text{whoever,} \\ \text{no one} \end{array} \right\} \Rightarrow W = \text{VERB}$$
15. 
$$\left. \begin{array}{l} \text{indef. Pr.} \\ \text{Obj.} \\ \text{it, you} \end{array} \right\} (+) W + \neq (N) \Rightarrow W = \text{VERB}$$

16.  $\neq (N) (+) W + \underline{itself} + V \Rightarrow W = \text{NOUN}$
17.  $W + \text{Refl. Pron.} \Rightarrow W = \text{VERB}$
18.  $W = \text{1st word in unit} \dots (V) = \left\{ \begin{array}{l} \text{aux.} \\ \text{modal} \\ \text{copula.} \end{array} \right\} \Rightarrow W = \text{ADJ.}$
19.  $\underline{as} + W + \left\{ \begin{array}{l} \text{punctuation} \\ \underline{as} \end{array} \right\} \Rightarrow W = \text{ADJ.}$   
 $\underline{as} + W + D \Rightarrow W = \text{ADJ.}$   
 $V = \begin{array}{l} \text{"appear,"} \\ \underline{\text{consider}} \end{array} + \dots + \underline{as} + W + \neq \left\{ \begin{array}{l} \text{adverb, } \underline{as} \\ \text{punctuation} \end{array} \right\} \Rightarrow W = \text{ADJ.}$
20.  $W (+) \left\{ \begin{array}{l} \text{Poss. A.} \\ \underline{\text{Whose}} \end{array} \right\} \Rightarrow W = \text{VERB}$
21.  $W = \underline{\text{consisting, etc.}} (+) \underline{\text{of}} \Rightarrow W = \text{VERB}$   
 $W \neq \underline{\text{consisting, etc.}} (+) \underline{\text{of}} \Rightarrow W = \text{ADJ.}$
22.  $\neq \text{Art.} + * \text{Sing. Count. N} + W \Rightarrow W = \text{NOUN}$
23.  $W + \underline{\text{due to}} \Rightarrow W = \text{NOUN}$
24.  $W (+) \text{Prep} \neq \underline{\text{to}} \Rightarrow W = \text{VERB}$
25.  $\left. \begin{array}{l} \underline{\text{most}} \\ \underline{\text{too}} \end{array} \right\} + W \Rightarrow W = \text{ADJ.}$
26.  $W + \text{Sing. Count. (N)} \Rightarrow W = \text{ADJ.}$
27.  $V + W \Rightarrow W = \text{ADJ.}$
28.  $\left. \begin{array}{l} \text{poss. N} \\ \text{poss. A} \end{array} \right\} + W \Rightarrow W = \text{ADJ.}$
29.  $W = \text{intr. (V)} + \left\{ \begin{array}{l} \text{N} \\ \text{A} \end{array} \right\} \Rightarrow W = \text{ADJ.}$   
 $W = \text{intr. (V)} + \neq \left\{ \begin{array}{l} \text{N} \\ \text{A} \end{array} \right\} \Rightarrow W = \text{VERB}$
30.  $\underline{\text{accident, etc.}} + \underline{\text{of}} (+) W \Rightarrow W = \text{VERB}$   
 $\neq \underline{\text{accident, etc.}} + \underline{\text{of}} (+) W \Rightarrow W = \text{ADJ.}$
31.  $\underline{\text{with, between}} (+) W \Rightarrow W = \text{ADJ.}$
32.  $\underline{\text{cure, etc.}} + \underline{\text{for}} (+) W \Rightarrow W = \text{VERB}$
33.  $\underline{\text{begin, etc.}} (+) \underline{\text{by}} (+) W \Rightarrow W = \text{VERB}$



34. Prep. (+) W + Punct.  $\Rightarrow$  W = ADJ.

Prep. (+) W + D  $\Rightarrow$  W = VERB

Prep. (+) W +  $\neq \begin{Bmatrix} \text{Pron.} \\ A \\ N \end{Bmatrix} \Rightarrow$  W = ADJ.

Prep. (+) W +  $\begin{Bmatrix} \text{Pron.} \\ A \\ N \end{Bmatrix} + \text{Prep} \Rightarrow$  W = VERB

Verb (+) Prep. (+) W + (Pron., A, N)  $\Rightarrow$  W = ADJ.

35. W + (N) (+) V  $\Rightarrow$  W = ADJ.

36. Let (+) W  $\Rightarrow$  W = ADJ.

37. Pron/A (+) W  $\Rightarrow$  W = ADJ.

38. A + Conj. or Punct. (+) W + N  $\Rightarrow$  W = ADJ.

39. = ing (+) conj. or punct. (+) W  $\Rightarrow$  W = Previous ing

40. V = begin etc. + . . . + W + Punct.  $\Rightarrow$  W = VERB

V  $\neq$  begin etc. + . . . + W + Punct.  $\Rightarrow$  W = ADJ.

41. W = VERB

APPENDIX XIV (cont'd.)

Past Tense Adjective Ambiguities Rules

NOTE: (P) = previous word, ignoring adverbs and adverb/adjectives  
F = following word  
Etc.

1. Is W an auxiliary, modal, or copulative? If yes, is (P) an auxiliary, modal, or copulative? If yes, take W as ADJ.; if no or if (P) is punctuation, is W auxiliary or modal? If yes, take as VERB.
2. Is (P) an article? If yes, take as ADJ.
3. Is (F) an article? If yes, take as VERB.
4. Is (P) a possessive adjective or possessive noun, or an adjective (ing)? If yes, take as ADJ.
5. Is (F) a possessive noun, possessive pronoun, possessive adjective, or whose? If yes, take as VERB.
6. Is (P) = very? If yes, take as ADJ.
7. Is (P) an auxiliary, copulative, or modal? If yes, take as ADJ.
8. Is (P) an indefinite or subject pronoun, whoever, who, no one, it, you? If yes, take as VERB.
9. Is (P) a preposition as? If yes, take as ADJ.
10. Is (F) an indefinite or object pronoun, no one, you, it? If yes, take as VERB.
11. Is (F) a reflexive pronoun? If yes, take as VERB.
12. Is (F) a preposition? If yes, take as VERB.
13. Is (W) the first word in the unit? If yes, is F an adjective, noun, subject or indefinite pronoun, it, or you? If yes, take as ADJ.
14. Is F a candidate noun? If yes, is (F)F a verb? If yes, take as ADJ.
15. Is F a singular countable noun? If yes, is FF a noun candidate. If no, or if FF is punctuation, take as ADJ.
16. Is (P) a noun, object pronoun, you, or it? If yes, is (P) (P) a verb auxiliary, 05, 08, or 17? If yes, take as ADJ.; if no, see if (P)(P) is a two object verb; if yes, is F a candidate noun? If yes, take as ADJ.
17. Is W capitalized but not the first word in the sentence? If yes, take W as NOUN.

17. Is (P) a noun? If yes, take as VERB.
18. Is (P) = as? If yes, is F = as? If yes, take as ADJ.
19. Is (P) = although, because, since, though, when, how, however, once, so, thus, what, whatever, wherever, that, as? If yes, is F punctuation? If yes, take as VERB.
20. Is (P) a verb? If yes, is F a candidate noun or an adjective? If yes take as ADJ.
21. Is (P) a pronoun/adjective? If yes, is (P)(P) a verb, preposition, or possessive adjective? If yes, take as ADJ.
22. Is (P) = although, because, since, when, though, how, if, until, where, whereas, whether, while, whilst, what? If yes, is F a candidate noun? If yes, take as ADJ.
23. Is (F) an English numeral, Arabic numeral, or other symbol? If yes, take as VERB.
24. Is (P) = that or which? If yes, is F an adjective/pronoun or adverb or punctuation? If yes, take as VERB; if no, take as ADJ.
25. Take W as VERB.

Past Tense Adjective Ambiguities Pictures

1.  $\left. \begin{array}{l} \text{auxiliary} \\ \text{modal} \\ \text{copulative} \end{array} \right\} \quad (+) W = \left\{ \begin{array}{l} \text{auxiliary} \\ \text{modal} \\ \text{copulative} \end{array} \right\} \Rightarrow W = \text{ADJ.}$
- $\neq \left\{ \begin{array}{l} \text{auxiliary} \\ \text{modal} \\ \text{copulative} \\ \text{punc.} \end{array} \right\} \quad (+) W = \left\{ \begin{array}{l} \text{auxiliary} \\ \text{modal} \end{array} \right\} \Rightarrow W = \text{VERB}$
2. article  $(+) W \Rightarrow W = \text{ADJ.}$
3. W  $(+)$  article  $\Rightarrow W = \text{VERB}$
4.  $\left. \begin{array}{l} \text{possessive noun} \\ \text{possessive adj.} \\ \text{adjective } (\neq \text{ing}) \end{array} \right\} \quad (+) W \Rightarrow W = \text{ADJ.}$
5.  $W (+) \left\{ \begin{array}{l} \text{poss. noun} \\ \text{poss. adj.} \\ \text{poss. pro.} \\ \text{whose} \end{array} \right\} \Rightarrow W = \text{VERB.}$
6. very  $(+) W \Rightarrow W = \text{ADJ.}$
7.  $\left. \begin{array}{l} \text{aux.} \\ \text{mod.} \\ \text{cop.} \end{array} \right\} \quad (+) W \Rightarrow W = \text{ADJ.}$
8.  $\left. \begin{array}{l} \text{indef. or subj. Pr.} \\ \text{who, whoever, it} \\ \text{you, no one} \end{array} \right\} \quad (+) W \Rightarrow W = \text{VERB}$
9. P  $\neq$  as  $(+) W \Rightarrow W = \text{ADJ.}$
10.  $W (+) \left\{ \begin{array}{l} \text{indef. or obj. Pr.} \\ \text{it, you, no one} \end{array} \right\} \Rightarrow W = \text{VERB}$
11. W  $(+)$  reflexive Pr.  $\Rightarrow W = \text{VERB.}$
12. W  $(+)$  preposition  $\Rightarrow W = \text{VERB.}$
13.  $\left( \begin{array}{l} \cdot \end{array} \right) \quad (+) W + \left\{ \begin{array}{l} \text{indef. or subj. Pr. A.} \\ \text{N., it, you} \end{array} \right\} \Rightarrow W = \text{ADJ.}$
14. W  $+$  (N)  $+$  V  $\Rightarrow W = \text{ADJ.}$
15. W  $+$  sing. count. N.  $+$  ( $\neq$  (N) )  $\Rightarrow W = \text{ADJ.}$

16.  $V = \left\{ \begin{array}{l} \text{subclass} \\ 05, 08, 17 \end{array} \right\} \quad (+) \quad \left\{ \begin{array}{l} \text{Object Pr.} \\ N, \underline{\text{you}}, \underline{\text{it}} \end{array} \right\} \quad (+) W \Rightarrow W = \text{ADJ.}$   
 two object V (+)  $\left\{ \begin{array}{l} \text{object pr.} \\ N, \underline{\text{you}}, \underline{\text{it}} \end{array} \right\} (+) W + (N) \Rightarrow W = \text{ADJ.}$
17.  $N (+) W \Rightarrow W = \text{VERB}$
18.  $\underline{\text{as}} (+) W + \underline{\text{as}} \Rightarrow W = \text{ADJ.}$
19.  $\left\{ \begin{array}{l} \underline{\text{although}}, \underline{\text{because}}, \underline{\text{since}}, \underline{\text{though}}, \\ \underline{\text{when}}, \underline{\text{how}}, \underline{\text{however}}, \underline{\text{once}}, \underline{\text{so}}, \underline{\text{there}}, \\ \underline{\text{whatever}}, \underline{\text{wherever}}, \underline{\text{that}}, \underline{\text{as}} \end{array} \right\} \quad (+) W + \text{punctuation} \Rightarrow W = \text{VERB.}$
20.  $V. (+) W + \left\{ \begin{array}{l} A \\ (N) \end{array} \right\} \Rightarrow W = \text{ADJ.}$
21.  $\left\{ \begin{array}{l} P, \\ \text{poss. A.} \end{array} \right\} \quad (+) \text{Pr/A} (+) W \Rightarrow W = \text{ADJ.}$
22.  $\left\{ \begin{array}{l} \underline{\text{although}}, \underline{\text{because}}, \underline{\text{since}}, \underline{\text{when}}, \underline{\text{though}}, \\ \underline{\text{how}}, \underline{\text{if}}, \underline{\text{until}}, \underline{\text{where}}, \underline{\text{whereas}}, \\ \underline{\text{whether}}, \underline{\text{while}}, \underline{\text{whilst}}, \underline{\text{what}} \end{array} \right\} \quad (+) W + (N) \Rightarrow W = \text{ADJ.}$
23.  $W (+) \left\{ \begin{array}{l} \text{English numeral} \\ \text{Arabic no.} \\ \text{symbol} \end{array} \right\} \Rightarrow W = \text{VERB.}$
24.  $\left\{ \begin{array}{l} \underline{\text{that}} \\ \underline{\text{which}} \end{array} \right\} \quad (+) W + \left\{ \begin{array}{l} \text{Pr/A} \\ D \\ \text{punctuation} \end{array} \right\} \Rightarrow W = \text{VERB}$
- $\left\{ \begin{array}{l} \underline{\text{that}} \\ \underline{\text{which}} \end{array} \right\} \quad (+) W + \left\{ \begin{array}{l} \text{Pr/A} \\ D \\ \text{punctuation} \end{array} \right\} \Rightarrow W = \text{ADJ.}$

APPENDIX XIV (cont'd.)

Present Tense Verb/Adjective Rules

NOTE: F(F) = word following (F)  
(F) = word following W, ignoring adverbs and adverb/adjectives  
F\* = same as (F), but also ignoring Adjectives (articles, possessives)

1. Is W capitalized but not the first word in the sentence? If yes, take W as NOUN.
2. Is W a + s (i.e. does it have an s affix)? If yes, take as VERB.
3. Is (P) an article, possessive noun, or possessive adjective? If yes, take as ADJ.
4. Is P = comparatively, too, strikingly, very? If yes, take as ADJ.
5. Is (P) = although, because, since, though, when, how, if, until, where, whereas, whether, while, whilst, what? If yes, take as ADJ.
6. Is (P) a subject pronoun? If yes, take as VERB.
7. Is (F) a possessive or reflexive or object pronoun, it, you? If yes, take as VERB.
8. Is (P) = another, any, each, much, no one, one, other, such, this? If yes, take as ADJ.
9. Is (P) a modal verb? If yes, take as VERB.
10. Is (P) an auxiliary or copulative verb? If yes, take as ADJ.
11. Is P = not? If yes, take as ADJ.
12. Is (P) a preposition (= to)? If yes, take as ADJ.
13. Is (P) a verb subclass 06, 25, 27? If yes, take as ADJ; if no, see if (F) is an article, possessive adjective, pronoun/adjective, indefinite pronoun, whoever, no one. If yes, is (P) a two object verb? If no, take as VERB.
14. Is F = of? If yes, take as ADJ.
15. Is P a pronoun/adjective, more, most? If yes, is (P)P a preposition, verb, or article? If yes, take as ADJ.
16. Is F an adverb? If yes, take as VERB.
17. Is (P) an intransitive verb? If yes, take as ADJ.
18. Is (P) an adjective? If yes, take as ADJ.

19. Is (P) an object or indefinite pronoun, no one, whoever, it, you?  
If yes, is (P)(P) = let? If yes, take as VERB.
20. Is (P) an indefinite pronoun, who, you? If yes, is (P)(P) a preposition or a verb? If no, take as VERB.
21. Is (P) a plural noun. If yes, is there a preposition or a verb preceding (P)? If no, take as VERB; if yes, is the Preceding Verb intransitive? If yes, take as VERB.
22. Is F a verb? If yes, take as NOUN.
23. Is W the first word in the unit? If yes, is F an Arabic numeral? If yes, take as VERB.
24. Is W the first word in the unit? If yes, is there a Following Verb candidate, F. (V.) in the unit? If no, take W as VERB; if yes, is F. (V.) a modal, copulative or auxiliary? If yes, take as ADJ.
25. Is F a candidate noun? If yes, is (F)F a verb? If yes, take as ADJ.
26. Is (P) a punctuation mark? If yes, is F a noun? If yes, is the Following Verb a modal, copulative or auxiliary? If yes, take as ADJ.
27. Is P = to? If yes, is PP = attributable, attribute(s)(ed), belong, opposed, opposite, proportion(al), similar, subject(ed)? If yes, take as ADJ.
28. Is (P) = to? If yes, is there a Preceding Verb in the unit? If no, take as VERB.
29. Is F\* a sing. count. noun candidate? If yes, is FF\* a noun candidate? If no, take as ADJ.
30. Is (P) a conjunction (i.e. and, or, nor, but, than) or punctuation? If yes, is P(P) an adjective? If yes, take as ADJ.
31. Is P = more, most? If yes, is PP a conjunction? If yes, is PPP an adjective? If yes, take W as ADJ.
32. Is (W) the first word in the unit? If yes, take as ADJ.
33. Is (P) punctuation? If yes, is (F) punctuation? If yes, is (P)(P) a verb? If yes, take as VERB; if no, take as ADJ.
34. Take W as VERB.

# Adjective/Present VERB Pictures

All searches start with W.

3. article  
poss. Noun  
poss. Adj. } (+) W  $\Rightarrow$  W = Adjective
4. comparatively, too,  
strikingly, very } + W  $\Rightarrow$  W = ADJ.
5. (although, . . . , how, . . . , what) (+) W  $\Rightarrow$  W = ATJ.
6. subject Pron. (+) W  $\Rightarrow$  W = VERB
7. W (+) { poss. Pro.  
reflex. Pro.  
Object Pro.  
it, you }  $\Rightarrow$  W = VERB
8. (another, . . . , this) (+) W  $\Rightarrow$  W = ADJ.
9. modal V (+) W  $\Rightarrow$  W = VERB
10. auxiliary } (+) W  $\Rightarrow$  W = ADJ.  
copulative }
11. not + W  $\Rightarrow$  W = ADJ.
12. Prep (to) + W  $\Rightarrow$  W = ADJ.
13. (V subclass 06, 25, 27) (+) W  $\Rightarrow$  W = ADJ.
13.  $\neq$  two object V (+) W (+) { article  
poss. ADJ.  
Pr/adj.  
indef. Pr.  
whoever, no one }  $\Rightarrow$  W = VERB.
14. W + of  $\Rightarrow$  W = ADJ.
15. { P  
V  
art. } (+) { Pr/A  
more  
most } (+) W  $\Rightarrow$  W = ADJ.
16. W + adverb  $\Rightarrow$  W = VERB.
17. intrans. V (+) W  $\Rightarrow$  W = ADJ.
18. A (+) W  $\Rightarrow$  W = ADJ.



19. let (+) {obj. or indef. Pr.  
whoever, no one, it, you} + W  $\Rightarrow$  W = VERB
20.  $\left. \begin{matrix} P \\ V \end{matrix} \right\}$  (+) {indef. Pr.  
who, you} (+) W  $\Rightarrow$  W = VERB
21.  $\left. \begin{matrix} P \\ V \end{matrix} \right\}$  + . . . + plural N (+) W  $\Rightarrow$  W = VERB  
intrans. V + . . . + plural N (+) W  $\Rightarrow$  W = VERB
22. W + V  $\Rightarrow$  W = NOUN
23. (W = 1st word in unit) + Arabic numeral  $\Rightarrow$  W = VERB
24. (W = 1st word in unit) + . . . +  $\neq$  (V)  $\Rightarrow$  W = VERB.  
(W = 1st word in unit) + . . . +  $\left\{ \begin{matrix} \text{modal} \\ \text{cop.} \\ \text{aux.} \end{matrix} \right\}$   $\Rightarrow$  W = ADJ.
25. W + (N) (+) V  $\Rightarrow$  W = ADJ.
26. Punct. (+) W + N + . . . +  $\left\{ \begin{matrix} \text{modal} \\ \text{cop.} \\ \text{aux.} \end{matrix} \right\}$   $\Rightarrow$  W = ADJ.
27. (attribute, . . . , subjected) + to + W  $\Rightarrow$  W = ADJ.
28. no V + . . . + to (+) W  $\Rightarrow$  W = VERB.
29. W (+)\* sing. count. (N)  $\neq$  (N)  $\Rightarrow$  W = ADJ.
30. A +  $\left\{ \begin{matrix} C \\ \text{Punct.} \end{matrix} \right\}$  (+) W  $\Rightarrow$  W = ADJ.
31. A + C +  $\left\{ \begin{matrix} \text{more} \\ \text{most} \end{matrix} \right\}$  + W  $\Rightarrow$  W = ADJ.
32. V (+) Punct. (+) W (+) Punct.  $\Rightarrow$  W = VERB  
 $\neq$  V (+) Punct. (+) W (+) Punct.  $\Rightarrow$  W = ADJ.
33. W = VERB

Past Tense - Noun Ambiguities Rules

Note: (F) = Following word (ignoring adverbs + adverb/adjectives).

1. Is W capitalised and not the first word in the sentence? If yes, take W as NOUN.
2. Is (P) an article, <sup>(u)</sup>possessive noun, or possessive adjective? If yes, take as NOUN.
3. Is (P) indefinite or object pronoun, no one, it, you? If yes, is (P)(P) = let. If yes, take as VERB.
4. Is (F) an object pronoun? If yes, take as VERB.
5. Is (P) a subject or indefinite pronoun, who, whoever, it, you? If yes, take as VERB.
6. Is (P) a modal verb? If yes, is W a+s (i.e. does W have an s affix.)? If no, take as VERB; if yes, take as NOUN.
7. Is (P) an auxiliary verb? If yes, take as ADJECTIVE.
8. Is (P) a verb? If yes, take as NOUN.
9. Is (P) a preposition ( ≠ to)? If yes, take as NOUN.
10. Is P an adjective ( ≠ ing) (i.e. an adjective not resolved by -ing routine)? If yes, take as NOUN.
11. Is (F) a verb? If yes, take as NOUN.
12. Is W the first word in the unit? If yes, is W a + s? If yes, take as NOUN; if no, see if there is a Following Verb candidate, F. (V.). If no, take as VERB; if yes, see if F.(V.) is an auxiliary, modal, or copulative. If yes, take as NOUN.
13. Is W a+s? If yes, is (P) = to? If yes, take as NOUN.
14. Is W = thought, spoke? If no, is (F) = of? If yes, take as NOUN.
15. Is W a+s? If yes, is P = another, each, no one, one, this? If yes, take as VERB; if no, see if P is a pronoun/adjective, that? If yes, take as NOUN.
16. Is P a pronoun/adjective, that? If yes, is (P)P a preposition or a verb? If yes, take as NOUN.
17. Is (F) a possessive noun, possessive pronoun, possessive adjective, or whose? If yes, take as VERB.
18. Is F a singular countable noun? If yes, take as NOUN.

19. Is (F) = adj. If yes, is W = felt? If yes, as V. If no, is F(F) = ~~pre~~-  
osition? If yes, take as NOUN.
20. Is P a singular countable noun? If yes, is (P)P punctuation, conjunction  
(i.e. and, or, nor, than, but), preposition, verb. If yes, take as NOUN;  
if no, is (P) an ing or past participial adjective? If yes, is (P)(P)P  
an auxiliary? If yes, take as NOUN.
21. Is W a+s? If yes, is (P) a noun, indefinite or object pronoun, you,  
it? If yes, is (P)(P) a two-object verb? If yes, take as NOUN.
22. Is (P) a plural noun? If yes, take as VERB.
23. Is W a+s? If yes, is P an Arabic or English numeral? If yes, take as NOUN.
24. If F an article, pronoun/adjective, indefinite pronoun, reflexive pronoun,  
more, most, very? If yes, take as VERB.
25. Is F a past tense verb candidate? If yes, is FF a noun or an adjective?  
If no, take as NOUN.
26. Is P an -ing word? If yes, take W as NOUN.
27. Is P punctuation? If yes, is F punctuation? If yes, is FF a verb or  
adjective? If no, take as NOUN; if yes, take W as PP.
28. Take W as VERB.

# Past Tense-Noun Ambiguities Pictures

1. W = cap ( ≠ 1st word in sent.)  $\Rightarrow$  W = NOUN
2.  $\left. \begin{array}{l} \text{article} \\ \text{possessive noun} \\ \text{possessive adjective} \end{array} \right\} (+) W \Rightarrow W = \text{NOUN}$
3. Let (+)  $\left\{ \begin{array}{l} \text{indefinite pronoun} \\ \text{object pronoun} \\ \underline{\text{no one}} \\ \underline{\text{it}} \\ \underline{\text{you}} \end{array} \right\} (+) W \Rightarrow W = \text{VERB}$
4. W (+) object pronoun  $\Rightarrow$  W = VERB
5.  $\left\{ \begin{array}{l} \text{subject pronoun} \\ \text{indefinite pronoun} \\ \underline{\text{who}} \\ \underline{\text{whoever}} \\ \underline{\text{it}} \\ \underline{\text{you}} \end{array} \right\} (+) W \Rightarrow W = \text{VERB}$
6. modal verb (+) W with s affix  $\Rightarrow$  W = NOUN  
modal verb (+) W with no s affix  $\Rightarrow$  W = VERB
7. auxiliary (+) W  $\Rightarrow$  W = ADJ.
8. verb (+) W  $\Rightarrow$  W = NOUN
9. preposition (≠ to) (+) W  $\Rightarrow$  W = NOUN
10. adjective with no -ing affix + W  $\Rightarrow$  W = NOUN
11. W (+) verb  $\Rightarrow$  W = NOUN
12. W = 1st word in unit and also with s affix  $\Rightarrow$  W = NOUN  
W = 1st word in unit + no FV  $\Rightarrow$  W = VERB  
W = 1st word in unit + FV =  $\left\{ \begin{array}{l} \text{auxiliary} \\ \text{modal} \\ \text{copulative} \end{array} \right\} \Rightarrow W = \text{NOUN}$
13. to (+) W with s affix  $\Rightarrow$  W = NOUN
14. W ≠  $\left\{ \begin{array}{l} \underline{\text{thought}} \\ \underline{\text{spoke}} \end{array} \right\} (+) \underline{\text{of}} \Rightarrow W = \text{NOUN}$
15.  $\left\{ \begin{array}{l} \underline{\text{another}} \\ \underline{\text{each}} \\ \underline{\text{no one}} \\ \underline{\text{one}} \\ \underline{\text{this}} \end{array} \right\} + W \text{ with } \underline{s} \text{ affix } \Rightarrow W = \text{VERB}$

15. (Con't.)  
pronoun/adjective + W with s affix  $\Rightarrow$  W = NOUN
16.  $\left. \begin{array}{l} \text{preposition} \\ \text{verb} \end{array} \right\} (+) \left\{ \begin{array}{l} \text{pronoun/adjective} \\ \text{that} \end{array} \right\} + W \Rightarrow W = \text{NOUN}$
17.  $W(+) \left\{ \begin{array}{l} \text{possessive noun} \\ \text{possessive pronoun} \\ \text{possessive adjective} \\ \text{whose} \end{array} \right\} \Rightarrow W = \text{NOUN}$
18. W = singular countable noun  $\Rightarrow$  W = NOUN
19. W ( $\neq$  felt) (+) adjective + preposition  $\Rightarrow$  W = NOUN
20.  $\left. \begin{array}{l} \text{punctuation} \\ \text{conjunction} \\ \text{(i.e. and, or,} \\ \text{nor, than, but} \\ \text{preposition} \\ \text{verb} \end{array} \right\} (\S) \left\{ \begin{array}{l} \text{singular} \\ \text{countable} \\ \text{noun} \end{array} \right\} + W \Rightarrow W = \text{NOUN}$
21. auxiliary (+) -ing or past participial A (+)  $\left\{ \begin{array}{l} \text{singular} \\ \text{countable} \\ \text{noun} \end{array} \right\} + W \Rightarrow W = \text{NOUN}$
21.  $\text{two object verb } (+) \left\{ \begin{array}{l} \text{noun} \\ \text{indefinite pronoun} \\ \text{object pronoun} \\ \text{you} \\ \text{it} \end{array} \right\} (+) W \text{ with } \underline{s} \text{ affix} \Rightarrow W = \text{NOUN}$
22. plural noun (+) W  $\Rightarrow$  W = VERB
23.  $\left. \begin{array}{l} \text{Arabic numeral} \\ \text{English numeral} \end{array} \right\} + W \text{ with } \underline{s} \text{ affix} \Rightarrow W = \text{NOUN}$
24.  $W + \left\{ \begin{array}{l} \text{article} \\ \text{pronoun/adjective} \\ \text{indefinite pronoun} \\ \text{reflexive pronoun} \\ \text{more} \\ \text{most} \\ \text{very} \end{array} \right\} \Rightarrow W = \text{VERB}$
25. W + past tense verb candidate +  $\neq \left\{ \begin{array}{l} \text{noun} \\ \text{adjective} \end{array} \right\} \Rightarrow W = \text{NOUN}$
26. -ing affix + W  $\Rightarrow$  W = NOUN
27. verb + punctuation + W + punctuation  $\Rightarrow$  W = VERB  
adjective + punctuation + W + punctuation  $\Rightarrow$  W = ADJECTIVE  
punctuation + W + punctuation  $\Rightarrow$  W = NOUN
28. W = VERB

## APPENDIX XIV (contd.)

### Idiosyncratic Distribution Rules

#### Notes Concerning Symbolism:

1. W = word under consideration  
P = word preceding W.  
F = word following W  
(P) = word preceding W, ignoring adverbs, and adverb/adjectives  
(F)\* = word following W, ignoring adverbs, adverb/adjectives, and adjectives (other than articles)  
PP = word preceding P  
P(P) = word preceding (P)
  2. These rules also set the proper form class code.
  3. When the location of the next instruction is not explicitly stated, the absence of this statement stands for the command: Go to the first step of the next rule.
- 

#### RULES

0. The following words and only the following words enter the odd ball routine for resolution: august; even; still, well; down, like, except, near, till; back; can, may, might, will; mine.
1. Is W capitalized and not the first word in the sentence? If yes, take W as NOUN.
2. Is W = august? If yes, is W capitalized? If yes, take W as a NOUN; if no, take W as an ADJECTIVE.
3. Is (P)\* an article? If yes, is (F) a preposition or a present tense verb + s? If yes, take W as a NOUN.
4. Is W = even? If yes, take W as an ADVERB.
5. Is (F) = a past tense verb candidate? If yes, is W = down, near, still, or well? If no, is (P)\* an article? If yes, take W as a NOUN.
6. 1) Is W = like or except? 2) If yes, is P an article or a possessive noun or possessive adjective? If yes, take W as a NOUN. 3) If no, is P = to? If no, go to question 5 of this rule. 4) If yes, is PP = verb (≠ list 20)? If yes, take W as a VERB. 5) If no, is (P) a modal, i. we, you, who, or they? If yes, take W as a VERB.
7. Is W = still or well? If yes, take W as an ADVERB.
8. Is P = off? If yes, take W as a NOUN.

9. Is W = down, like, except, near or till? If yes, take W as a PREPOSITION.
10. Is W = back? If yes, take W as an ADVERB.
11. Is W = mine? If no, take W as a VERB.
12. Is P a preposition (to)? If yes, take W as a possessive PRONOUN. If no, is P = to? If yes, is PP a member of list 20 or 21? If yes, take W as a possessive PRONOUN.
13. Is (P) a conjunction? If yes, is P(P) an adjective or possessive pronoun? If yes, take W as a possessive PRONOUN.
14. Set the NOUN/VERB form class codes.

## APPENDIX XV -- Absolute Breakers

### I.

because  
how  
if  
what  
when  
whenever  
where  
whereas  
wherever  
whether  
which  
while  
who  
whom  
whose  
why  
although  
whichever  
whoever  
whomever  
whosoever  
albeit  
unless  
though  
whereby  
insofar  
whereat  
wherein  
whereafter  
wherefore  
lest  
whilst

V-ing

whensoever  
whence  
whencasoever  
wherefrom  
whereinto  
whereof  
whereon  
werethrough  
wereto  
whereunto  
whereupon  
werewith



## II.

at the same time as  
at the same time that  
as if  
as though  
by the time  
else that  
even if  
even though  
for fear that  
for then  
however many  
however much  
if and only if  
inasmuch as  
in order not to  
in order that  
in order to  
in so far as  
in the hope that  
just as if  
just as though  
just because  
not even if  
not even though  
now that  
or else  
so as  
so that  
such that  
the way in which  
the way that

### III - 2

(1)				
('provided)	+	<u>that</u>	(+)	article
" "	"	"	"	subject pronoun
" "	"	"	"	indefinite pronoun
" "	"	"	"	possessive pronoun
" "	"	"	"	possessive adjective
" "	"	"	"	plural noun
" "	"	"	"	proper name
" "	"	"	"	verb
" "	"	"	"	<u>more</u>
" "	"	"	"	<u>most</u>
" "	"	"	"	<u>that</u>
" "	"	"	"	adjective /+ / plural noun

'Conjunction (and, but, or, nor, or, than) (+) verb

Preposition (to) (+) verb

'to + present tense verb with no s affix

to (+) 'verb (not a singular present tense verb)

irregular verbal past participle (driven, lain, etc.)

'after (+) past tense verb or past participle

'until " " " " " " "

'since " " " " " " "

'before " " " " " " "

verb (+) ' conjunction

'as + it

" " " subject pronoun

" " " verb

'so + article

" " " noun

" " " pronoun

" " " verb

'until + it, you

'since + it, you

'however + adverb/adjective

punctuation + 'but + subject pronoun  
" " " + proper name

'after + subject pronoun

'until " " "

'since " " "

'before " " "

'besides " " "

object pronoun	+	:	subject pronoun
indefinite pronoun	"	"	"
preposition	"	"	"
adjective/pronoun	"	"	"
noun	"	"	"
verb ≠ (modal, aux.)	"	"	"
indefinite pronoun	(+)	:	<u>that</u>
possessive pronoun	"	"	"
possessive adjective	"	"	"
dictionary adjective	"	"	"
present tense adjective	•	"	"
<u>such</u>	"	"	"
<u>sense</u>	"	"	"
<u>reason</u>	"	"	"

# APPENDIX XVI

## - Clause Division Rules -

Conventions: / means clause break; + followed by; (+) followed by, ignoring adverbs and adjective-adverbs; /+ followed by, ignoring everything except Verbs and Punctuation; /+\* followed by, ignoring everything except Verbs. NG means Nominal group; PG nominal group headed by a preposition. V means Verb; BV "believe" -- type Verb. # means that the item so marked if present in the specified position makes the rule inoperative. T means transitive, I intransitive. # means beginning of sentence.

### Rule 1

absolute breaker list 1 /+ /, /+ V  
absolute breaker list 1 /+ /, /+ # /+ V

### Rule 2

absolute breaker list 1 (+) abs. br. list 1 (+) Adjective (+)  
/NG # {plural N  
uncount. N} /+\* V

### Rule 2a

abs. br. list 1 (+) abs. br. list 1 (+) adjective + # {plural N  
uncount N} /+\*/V

### Rule 3

/, /+ V (+) /, (+) V

### Rule 4

/, (+) ed/en- V (+) # NG /+ /, (+) V

### Rule 5

V # (BV (+) that + prep) /+ [ # adjective (+) /, (+) # adjective  
# adverb + /, + # adverb  
# noun (+) /, + # {NG  
NG + PG} + {or  
and} ] /+ V

### Rule 6

V /+ /, NG, /+/, (+) V # (T-ed (+) # NG)

### Rule 7

V /+\* V T-ed (+) # NG

### Rule 8

V # 2 object V /+\* {noun  
indefinite pronoun  
reflexive pronoun  
possessive pronoun} (+) / that /+\* V

Rule 8a

V ≠ 2 object V /+/\* {noun  
indefinite pronoun  
reflexive pronoun  
possessive pronoun} (+), +/that /+/\* V

Rule 9

V ≠ 2-object V /+/\* {noun  
indef. pro.  
reflex. pro.  
poss. pro.} (+) / {article  
subject pro.  
indef. pro.  
poss. adj.  
noun (+) / {pronoun-adjective  
adjective + ≠ preposition  
plural noun (+) / noun  
indefinite article + noun (+) / plural noun  
adjective ≠ comparative (+) / {definite article  
subject pro.  
poss. adj.  
≠ so (+) adjective (+) / indefinite article} /+/\* V

Rule 9a

V /+/\* {noun  
indef. pro.  
reflex. pro.  
poss. pro.} (+) /, (+) {article  
sub. pro.  
indef. pro.  
poss. adj.  
noun (+) /, (+) {pro. adj.  
adj. + ≠ prep.  
plural noun (+) /, (+) noun  
indef. article + noun (+) /, (+) plural noun  
adj. ≠ comparative (+) /, {(+)  
definite article  
sub. pro.  
poss. adj.  
≠ so (+) adj. (+) /, (+) indef. article} /+/\* V

Rule 10

auxiliary } BV  
copulative } (+) BV adj.  
it (+) {copulative  
auxiliary (+) copulative adj.  
happens  
aux. (+) happened/ing} /+/\* {/ there  
≠ Prep + / that} /+/\* V

Rule 11

any form of  
be (+) / that /+/\* V

Rule 12 V / +/\* so (+) ing/ed adj. (+)/that /+/\* V

Rule 13 { adjective comparative } /+/\* V /+/\* /the + { adj. comparative } /+/\* V  
the + { more }

Rule 14  
 # (+) { if } /+/\* V /+/\* /then /+/\* V  
had

Rule 15

V /+/\* / { than  
as + { adverb } + as  
just as  
so + { long } + as  
as + to } /+/\* V

Rule 16  
 aux. + I <sup>VI</sup> adj. } ≠ copulative (+)/NG /+/\* V

Rule 17  
 # { NG } (+) { who  
which  
that  
whose (+) NG } (+) V /+/\* /V

Rule 18  
 auxiliary } <sup>BV<sup>I</sup></sup>  
 copulative } (+) BV-adjective I } /+/\* V

Rule 19

noun (+) { subj. pro  
article  
indef. pro.  
poss. pro.  
poss. adj.  
adj. } (+) V ≠ { 2 object } /+/\* /NG /+/\* V  
BV  
 plural noun (+) noun

Rule 20

it (+) { to + BV  
 aux. (+) BV-adj.  
 copulative  
 aux. + copulative-adj.  
happens  
 aux. (+) happened/ing } /+/\* V

# Rule 21

Prep. +  $\left\{ \begin{array}{l} \text{that} \\ \text{what} \\ \text{who} \\ \text{whose} \\ \text{whom} \\ \text{which} \\ \text{whether} \\ \text{why} \\ \text{how} \\ \text{where} \end{array} \right\} \quad /+/* \text{ V } /+/* \quad \underline{\text{there}} \quad (+) \quad \left\{ \begin{array}{l} \text{auxiliary} \\ \text{copulative} \\ \text{I} \\ \text{modal (+)} \left\{ \begin{array}{l} \text{aux.} \\ \text{cop.} \\ \text{I} \end{array} \right\} \end{array} \right\} \quad (+) \quad \left\{ \begin{array}{l} \text{NG} \\ \text{infinitive} \end{array} \right\}$

# Rule 21a

$\neq \left\{ \begin{array}{l} \text{that} \\ \text{what} \\ \text{who} \\ \text{whose} \\ \text{whom} \\ \text{which} \\ \text{whether} \\ \text{why} \\ \text{how} \\ \text{when} \end{array} \right\} \quad /+/* \text{ V } /+/* \quad \underline{\text{there}} \quad (+) \quad \left\{ \begin{array}{l} \text{aux.} \\ \text{cop.} \\ \text{I} \\ \text{modal (+)} \left\{ \begin{array}{l} \text{aux.} \\ \text{cop.} \\ \text{I} \end{array} \right\} \end{array} \right\} \quad (+) \quad \left\{ \begin{array}{l} \text{NG} \\ \text{infinitive} \end{array} \right\}$

# Rule 22

$\# + \left\{ \begin{array}{l} \text{that} + \text{article} \\ \text{that} + \text{plural noun} \\ \text{what} \\ \text{who} \\ \text{whose} \\ \text{whom} \\ \text{which} \\ \text{whether} \\ \text{why} \\ \text{how} \\ \text{where} \\ \text{when} \\ \text{the fact that} \\ \text{the reason that} \\ \text{the reason why} \end{array} \right\} \quad /+/* \text{ V } \quad /+/* \quad /V = \left\{ \begin{array}{l} \text{I} \\ \text{T} \\ \text{T/I} \end{array} \right\} \quad (+) \text{ NG} \}$

# Rule 23

$\left\{ \begin{array}{l} \text{V-en} \\ \text{V-ing} \\ \text{V-ed} = \left\{ \begin{array}{l} \text{T} \\ \text{T/I} \end{array} \right\} \end{array} \right\} \quad (+) \neq \text{NG} \quad /+/* \quad \left\{ \begin{array}{l} \text{auxiliary} \\ \text{modal} \\ \text{copulative} \\ \text{pres. tense} \\ \text{past tense only} \\ \text{V-ed} = \left\{ \begin{array}{l} \text{T} \\ \text{T/I} \end{array} \right\} \end{array} \right\} \quad (+) \text{NG}$

# Rule 24

Verb /+/\* PG +  $\neq \left\{ \begin{array}{l} \text{plural N} \\ \text{uncountable N} \end{array} \right\} \quad /+/* \quad /V$

Rule 25

Be (+) {Adj. = {-ed  
-en } ≠ 2 object } (+) / {NG  
that } /+/\* V

Rule 26

V pres. tense }  
auxiliary } /+/\* / { V pres. tense  
auxiliary  
modal  
copulative



APPENDIX VII (cont)

Sample Sentences for Clause Division Rules

- Rule 1      The whale, though not a fish, lives in the sea.
- Rule 2      They published a story which though inexact the secretary released to the newsmen.
- Rule 2a     The story which though inexact in all details was widely publicised aroused a great deal of interest.
- Rule 3      The Africans, Livingstone reported earlier, were friendly people.
- Rule 4      The man, utterly humiliated before his family, never returned.
- Rule 5      After they won the race, the team travelled to Europe.
- Rule 6      Due to perturbation caused by stars, such clouds, and the resulting clusters, probably never assumed a definite shape.
- Rule 7      They found stains caused by fire.
- Rule 8      They resent the idea that in many cases power wins over justice.
- Rule 8a     They resented the idea, as a matter of course, that power should win over justice.
- Rule 9      He said something nobody would believe.
- Rule 10     It was assumed that evidence could be found.
- Rule 11     The problem was that money was not appropriated.
- Rule 12     The house was so completely dilapidated that reconstruction was out of the question.
- Rule 13     The more it rains the greater the amount of plant growth will be.
- Rule 14     If Castro is a genius then I eat my hat.
- Rule 15     He worked as hard as he could.
- Rule 16     Wherever he went a big crowd turned out to cheer him.
- Rule 17     People who live in glass houses are nervous.
- Rule 18     He began to think the Cubans were insane.
- Rule 19     In the discussion we had most issues were settled.
- Rule 20     They pretend the missiles are being dismantled.

- Rule 21 On the question of whether the motion should be adopted there is total disagreement.
- Rule 21a In areas drained of water there is drought.
- Rule 22 That people are basically good follows from what was said about God.
- Rule 23 The method used by the first team was based on false premises.
- Rule 24 The sky when you look through a telescope seems nearer.
- Rule 25 When the work is done people tend to relax.
- Rule 26 The phases of the moon which influence our weather have been observed by farmers for many centuries.
- Rule 27 This seems, even to modern observers, extremely well constructed.

TH

APPENDIX XVII -- Randomly Selected Scientific Corpus

Agriculture

1. \*N.H. Winstead, J. C. Wells & W. W. Reid, BOTRYTIS IN PEPPER (CAPSICUM FRUTESCENS) SEEDBEDS & ON YOUNG PLANTS IN THE FIELD, Plant Dis. Reporter, Volume 42, No. 8, p. 981-982, August 15, 1958.
2. \*J. C. Robertson & L. E. Barhart, INTERRELATIONSHIPS BETWEEN DRIED CORN DISTILLERS SOLUBLES & CERTAIN MINERALS IN THE PREVENTION OF PARAKERATOSIS IN SWINE, (Abs) Animal Science, Vol. 17, No. 4, p. 1184, November, 1958.
3. MINNESOTA DEPARTMENT OF AGRICULTURE, DAIRY FOOD SECTION OF FEED & CONTROL, Fertilizer Analyses & Registrations 1958, St. Paul, 81 p.
4. \*E. H. McNally & A. W. Brant, OBSERVATIONS ON YOLK MOLDING, Poultry Science, Vol. 37, No. 5, p. 1225-6, September, 1958.
5. \*E. D. Goldsmith, EFFECTS OF COLCHICINE & DEMECOLCINE ON DEVELOPMENT OF THE FRUIT FLY, DROSOPHILA MELANOGASTER, (Abs) Anat. Rec., Vol. 131, No. 3, p. 558-9, September, 1958.
6. \*G. E. Templeton, EFFECT OF TOXIC METABOLITES FROM FUSARIUM MONILIFORME ON GERMINATION OF BARLEY, Dissertation Abstracts, Vol. 19, No. 7, p. 1535-6, January 1959.
7. G. L. Peterson, AGRICULTURAL ACTIVITIES OF U.S. OPERATIONS MISSION IN ISRAEL, Israel Econ. Forum, Vol. 9, No. 1, p. 40-64, December 1958.
8. G. De Lotto, THE PSEUDOCOCCIDAE (HOM:COCCOIDEA) DESCRIBED BY C. K. BRAIN FROM SOUTH AFRICA BRITISH MUSEUM (NATURAL HISTORY), British Entomology, Vol. 7, No. 3, p. 77-120, August 1958.
9. J. P. Latham, THE DISTANCE RELATIONS OF CROPLAND AREAS IN PENNSYLVANIA Annals of the American Association of Geography, Vol. 48, No. 3, p. 277, September 1958.
10. J. P. Cooper, THE GENETIC STRUCTURE OF RYEGRASS STRAINS AND ITS IMPLI-  
CATIONS FOR HERBAGE PLANT BREEDING, (Abs) Heredity, Vol. 12, No. 3, p. 400, August 1958.
11. L. A. Alban & C. J. Lin, EFFECTS OF LIME ADDITIONS ON pH & BASE SATURATION OF FIVE WESTERN OREGON SOILS, Soil Science, Vol. 86, No. 5, p. 271-5, November 1958.
12. D. H. Van Slogteren, SOME RECENT DEVELOPMENTS IN INDEXING VIRUS INFECTED POTATO PLANTS IN THE NETHERLANDS, (Abs) American Potato Journal, Vol. 36, No. 8, p. 303-4, August, 1958.
13. P. F. Griffin, URBAN IMPACT ON AGRICULTURE IN SANTA CLARA COUNTY, CALIF. (Abs) Assoc. American Geography. Annals of the American Association of Geography, Vol. 48, No. 3, p. 256, September, 1958.

14. E. T. Relf, DETERMINATION OF DRY SUBSTANCE IN MOLASSES & MASSECUITES, Queensland Society of Sugar Technology, Proceedings, No. 25, p. 189-93, 1958.
15. H. Ris, CHROMOSOME STRUCTURE. SYMPOSIUM, Chemical Basis of Heredity 1956 p. 23-69, Pub. 1957.
16. J. W. Bailey, 3 WAYS TO GET RID OF LICE AND MITES, Poultry Tribune, Vol. 65, No. 2, p. 62-3, February 1959.
17. I. Watson, 4-H LIVESTOCK JUDGING MANUAL: BEEF CATTLE-SHEEP-SWINE-DAIRY CATTLE, New Mexico Agriculture College, Extension C. No. 294, p. 26, March 1959.
18. W. Burniston, THE DESIGN & DEVELOPMENT OF THE ALUMINUM ALLOY MILK CAN, Dairy Industry, Vol. 23, No. 11, p. 826-9, November 1958.
19. Tefertiller, K., W. B. Black & L. Tweenen, MARKET & PRICES FOR OKLAHOMA BROOM CORN, Oklahoma Agricultural Experiment Station B, No. 23P, Dec. 1958.
20. J. R. Piening, REPORT ON MALT BEVERAGES, SIRUPS, EXTRACTS & BREWING MATERIALS, Assoc. Off Agr Chemistry, Vol. 42, No. 2, p. 330, May, 1959.
21. R. Schessler, THE FLAVOR OF VANILLA, Americas, Vol. 11, No. 8, p. 13-16, August 1959.
22. Institute of Food Technologist, WHAT TRAINING SHOULD A 4-YEAR FOOD TECHNOLOGY STUDENT RECEIVE?, SUMMARY REPORT ON THE EDUCATIONAL CONFERENCE MONTICELLO, ILL., MAY 1958, Food Technology, Vol. 12, No. 9, p. 7-8, 11, 13-14, September 1958.
23. R. R. Davies, DETACHMENT OF CONDIA BY CLOUD DROPLETS, Nature (London), Vol. 183, No. 4674, p. 1659, June 13, 1958.
24. R. Johnson, FRUIT GROWERS CAUGHT IN COST PRICE SQUEEZE FEAR FURTHER COST INCREASE, Sun Sweet Standard, Vol. 42, No. 11, p. 11, April 1959.
25. C. William, HOW DAIRYING IN INDIA CAN BE IMPROVED?, Madras, Vet. Col. Annual 17, p. 23-5, March 1959.

# APPENDIX Q (con't.)

## ASTRONOMY

1. \*G. Wallerstein, NOTE ON THE BEHAVIOR OF THE RV TAURI-TYPE STAR No. 11, IN MESSIER 2, Astronomical Journal, Vol. 62, p. 168, 1957.
2. \*W. Palmstorfer, AN AUSTRIAN AMATEUR PHOTOGRAPHS A BRIGHT METEOR, Sky & Telescope, Vol. 16, p. 301, April, 1957.
3. SYMPOSIUM ON RADIO ASTRONOMY, Commonwealth Scientific & Industrial Research Organization Radiophysics Laboratory, Sydney, September 1956; Melbourne, 1957.
4. \*I. King, THE DYNAMICAL LIFETIME OF A STAR CLUSTER, Astronomical Journal, Vol. 62, p. 144, 1957.
5. \*S. F. Singer, THE EFFECT OF METEORIC PARTICLES ON A SATELLITE, American Rocket Society, No. 307, p. 1-22, 1956.
6. \*H. M. Johnson, THE KINEMATICS & EVOLUTION OF POPULATION I STARS, Publications of the Astronomical Society of the Pacific, Vol. 69, p. 54-8, 1957.
7. Y. Nakagawa, EXPERIMENTS ON THE INHIBITION OF THERMAL CONVECTION BY A MAGNETIC FIELD, Proceedings of the Royal Society, Series A, Vol. 240, p. 108-113, 1957.
8. C. O. Hines & E. L. Vogan, VARIATIONS IN THE INTRINSIC STRENGTH OF THE 1956 QUADRANTID METEOR SHOWER, Canadian Journal of Physics, Vol. 35, p. 703-711, 1957.
9. E. J. Long, TRACKING THE SATELLITE, Nature Magazine, Vol. 50, No. 3, p. 154-5, 162, 1957.
10. G. Keller, STUDIES OF THE BEHAVIOR OF STELLAR SHADOW PATTERNS, Journal of the Optical Society of America, Vol. 47, p. 1047, 1957.
11. G. C. Scorgie, ON FREE MOTION IN THE GRAVITATIONAL FIELD OF THE EARTH, Quarterly Journal of Mechanics and Applied Mathematics, Vol. 10, p. 494-9, 1957.
12. E. Rabe, FURTHER STUDIES ON THE ORBITAL DEVELOPMENT OF PLUTO, Astrophysical Journal, Vol. 126, p. 240-44, 1957.
13. M. W. Mayall, VARIABLE STAR NOTES, AMERICAN ASSOCIATION OF VARIABLE STAR OBSERVERS, Journal of the Royal Astronomical Society of Canada, Vol. 51, p. 109-112, 165-8, 219-222, 259-62, 308-11, 353-6, 1957.
14. K. M. Yoss, OBJECTIVE-PRISM EQUIVALENT WIDTHS OF EARLY-TYPE STARS, Astronomical Journal, Vol. 62, p. 42, 1957.
15. D. B. McLaughlin, RECENT CLOSE APPROACH OF MARS, Science, Vol. 125, p. 893-894, 1957.

16. R. I. Mitchell, PHOTOELECTRIC OBSERVATIONS OF URANUS & NEPTUNE, Publication of the Astronomical Society of the Pacific, Vol. 69, p. 565-8, 1957.
17. W. P. Bidelman & A. McKellar, DOUBLE LINES IN THE SPECTRUM OF CASSIOPEIAE, Publications of the Astronomical Society of the Pacific, Vol. 69, p. 31-40, 1957.
18. J. T. Jefferies & F. Q. Orrall, THE INTERPRETATION OF BALMER LINE PROFILES IN SOLAR PROMINENCES, Astronomical Journal, Vol. 62, p. 143-4, 1957.
19. P. C. Keenan & J. A. Wright, ON THE SPECTRAL TYPE & LUMINOSITY OF VV CEPHEI, Publications of the Astronomical Society of the Pacific, Vol. 69, p. 457-8, 1957.
20. S. Chandrasekhar, THE THERMAL INSTABILITY OF A ROTATING FLUID SPHERE HEATED WITHIN II., Philosophical Magazine Series 8, Vol. 2, p. 1282-4, 1957.
21. X-RAY EMISSION DURING SOLAR FLARES, Sky & Telescope, Vol. 16, p. 223, 1957.
22. H. K. Eichhorn, SYSTEMATIC ERRORS IN THE HELSINGFORS ZONE OF THE CARTE DU CIEL, Astronomical Journal, Vol. 62, p. 142, 1957.
23. GEOPHYSICS SYMPOSIUM ON SOLAR & TERRESTRIAL RELATIONSHIPS, Proceedings of the National Academy of Sciences -U.S.A., Vol. 43, p. 1-104, 1957.
24. L. Herzberg, THE WAVELENGTH DISPLACEMENTS OF SOME INFRARED LINES BETWEEN LIMB & CENTER OF THE SUN, Canadian Journal of Physics, Vol. 35, p. 766-767, 1957.
25. L. A. Pars, INEQUALITIES OCCURRING IN THE RESTRICTED PROBLEM OF THREE BODIES, Journal of the London Mathematical Society, Vol. 32, p. 355-6, 1957.

## BIOLOGY

1. \*Hubert Frings & Mable Frings, NOTES ON REARING THE MOTHS, CISSEPS, (SCEPSIS) FULVICOLLIS (HBN) & CTENUCHA VIRGINIA, (CHARP.) IN THE LABORATORY, Journal of Economic Entomology, Vol. 52, No. 1, p. 168-9, 1959.
2. \*William Lassiter & John B. Stanbury, THE IN VIVO CONVERSION OF THYROXINE TO 3:5:3' -TRIIODOTHYRONINE, Journal of Clinical Endocrinology & Metabolism, Vol. 18, No. 8, p. 903-6, 1958.
3. \*Howard R. Engel, D. M. Bergenstal, Wilbert Nixon & James A. Patten, EFFECT OF HUMAN GROWTH HORMONE ON UFA & PLASMA AMINO ACID NITROGENS IN MAN, Proceedings of the Society for Experimental Biology and Medicine, Vol. 100, No. 4, p. 699-701, 1959.
4. \*John P. Fox, PROPHYLAXIS AGAINST RABIES IN HUMANS. In: ANIMAL DISEASE & HUMAN HEALTH, Annals of the New York Academy of Sciences, Vol. 70, No. 3,
5. \*Calvin McMillan, SURVIVAL OF TRANSPLANTED CUPRESSUS IN THE PYGMY FORESTS OF MENDOCINO COUNTY, CALIFORNIA, Madrono, Journal of the California Botanical Society, Vol. 15, No. 1, p. 1-4, 1959.
6. Daniel H. Funkenstein, NOR-EPINEPHRINE-LIKE AND EPINEPHRINE-LIKE SUBSTANCES IN RELATION TO HUMAN BEHAVIOR, Journal of Nervous & Mental Diseases, Vol. 124, No. 1, p. 58-68, 1956.
7. Robert L. Schultz, Edith A. Maynard & Daniel C. Pease, ELECTRON MICROSCOPY OF NEURONS, & NEUROGLIA OF CEREBRAL CORTEX & CORPUS CALLOSUM, American Journal of Anatomy, Vol. 100, No. 3, p. 369-407, 1959.
8. C. H. Waddington, COMMENT ON PROFESSOR STERN'S LETTER, American Naturalist, Vol. 92, No. 867, p. 375-6, 1958.
9. J. A. Wier, SEX RATIO RELATED TO SPERM SOURCE IN MICE, Journal of Heredity, Vol. 43, No. 2, P. 129-30, 1959.
10. C. E. Yarwood, VIRUS INFECTION & HEATING REDUCE SMOG DAMAGE, Plant Disease Reporter, Vol. 43, No. 2, p. 129-30, 1959.
11. Eva F. Estermann, G. H. Peterson & A. D. McLaren, DIGESTION OF CLAY-PROTEIN, LIGNIN-PROTEIN, & SILICA-PROTEIN COMPLEXES BY ENZYMES & BACTERIA, Soil Science Society of America, Proceedings, Vol. 23, No. 1, p. 31-6, 1959.
12. H. B. Barlow, TEMPORAL & SPATIAL SUMMATION IN HUMAN VISION AT DIFFERENT BACKGROUND INTENSITIES, Journal of Physiology, Vol. 141, No. 2, p. 337-50.
13. Harry Grundfest, GENERAL PHYSIOLOGY & PHARMACOLOGY OF SYNAPSES & SOME IMPLICATIONS FOR THE MAMMALIAN CENTRAL NERVOUS SYSTEM, Journal of Nervous & Mental Diseases, Vol. 128, No. 6, p. 473-496, 1959.

14. John Dunleavy, STUDIES OF A SEEDLING BLIGHT OF SOYBEANS & THE ETIOLOGY OF CAUSAL FUNGUS, DIAPORTHE PHASEOLORUM VAR. CAULIVORA., Proceedings of the Iowa Academy of Science, Vol. 65, p. 131-45, 1958.
15. Derek Goodwin, THE EXISTENCE & CAUSATION OF COLOUR-PREFERENCES IN THE PAIRING OF FERAL & DOMESTIC PIGEONS, Bulletin of the British Ornithologist's Club, Vol. 78, No. 8, p. 136-9, 1958.
16. Peter B. Wright & Louis P. Brady, AN ANATOMIC EVALUATION OF WHIPLASH INJURIES, Clinical Orthopaedics, Vol. 11, p. 120-131, 1958.
17. M. B. Mitchell, GENETIC RECOMBINATION IN NEUROSPORA, Genetics, Vol. 43, No. 6, p. 799-813, 1958.
18. Nolan E. Rice, OCCURRENCE OF GLASS SPONGES IN THE DEVONIAN OF WEST VIRGINIA, Journal of Paleontology, Vol. 32, No. 1, p. 244-5, 1958.
19. D. A. Lindquist, M. L. Fairchild, P. A. Dahm & J. Gurland, THIODAN RESIDUES ON CORN PLANTS, Journal of Economic Entomology, Vol. 52, No. 1, p. 102-6, 1959.
20. Raymond E. Knauff, Stefan S. Fajans, Eduardo Ramirez & Jerome W. Conn, METABOLIC STUDIES ON CHLORPROPAMIDE IN NORMAL MEN & IN DIABETIC SUBJECTS. In: CHLORPROPAMIDE & DIABETES MELLITUS, Annals of the New York Academy of Sciences, Vol. 74, No. 3, p. 603-17, 1959.
21. John H. Annegers, ABSORPTION OF LIPID FOLLOWING THORACIC DUCT LIGATION IN THE RAT, American Journal of Physiology, Vol. 196, No. 2, p. 363-4, 1959.
22. Hava Stein, A CORRELATION BETWEEN ANTHER PIGMENTATION & POLLEN STERILITY IN BETA VULGARIS, Dissertation Abstracts, Vol. 18, No. 3, p. 752, 1958.
23. Henry Van der Schalie, PROBLEMS IN THE CONTROL OF INTERMEDIATE HOSTS (ONCOMELANIA & POMATIOPSIS) OF THE ORIENTAL BLOOD FLÜKE, American Malacological Union, Annual Report, 1957, p. 12-13, 1957.
24. D. Keilin, THE LEEUWENHOEK LECTURE. THE PROBLEM OF ANAVIOSIS OR LATENT LIFE: HISTORY AND CURRENT CONCEPT, Proceedings of the Royal Society, Series B, Biological Sciences, Vol. 150, No. 939, p. 149-91, 1959.
25. F. John Vernberg & David K. Meriney, THE INFLUENCE OF TEMPERATURE & HUMIDITY ON THE METABOLISM OF MELANISTIC STRAINS OF DROSOPHILA MELANOGASTER, Journal of the Elisha Mitchell Scientific Society, Vol. 73, No. 2, p. 351-62, 1957.



CHEMISTRY

1. \*John W. Lynn,  $\alpha, \alpha$ -DIMETHYL- $\beta$ -HYDROXYPROIONALDEHYDE AND  $\alpha$ -HYDROXY- $\beta$ ,  $\beta$ -DIMETHYL- $\gamma$ -BUTYROLACTONE. (to Union Carbide Corporation). U.S. 2,836,878, December 9, 1958.
2. \*A. W. Weitkamp, ANALYTICAL SEPARATIONS IN THE STUDY OF LIMONENE SULFIDES. The Perfumery and Essential Oil Record, Vol. 49, pp. 803-7, 1958.
3. Sigmund Goldstein, REDUCTIC ACID PREPARATION FROM PLANT MATERIALS. U.S. 2,854,484, September 30, 1958.
4. \*Howard R. Guest, and Harry A. Stansbury Jr., HALOGENATED DIEPOXIDE COMPOUNDS. (To Union Carbide Corporation). U.S. 2,874,167, Feb. 17, 1959.
5. \*Henry Gilman & Suzanne Gray, METALATION REACTIONS IN TETRAHYDROFURAN, Journal of Organic Chemistry, Vol. 23, pp. 1476-9, 1958.
6. \*Carl D. Thurmond, SEMICONDUCTOR CONTACTS. (to Bell Telephone Laboratories Inc.) U.S. 2,877,147, March 10, 1959.
7. E. A. Delwiche, G. M. Fukui, A. W. Andrews, & M. J. Surgalla. ENVIRONMENTAL CONDITIONS AFFECTING THE POPULATION DYNAMICS & THE RETENTION OF VIRULENCE OF PASTEURILLA PESTIS: THE ROLE OF CARBON DIOXIDE, Journal of Bacteriology, Vol. 77, pp. 355-360, 1959.
8. J. W. Millen & A. Hess, THE BLOOD BRAIN BARRIER, AN EXPERIMENTAL STUDY WITH VITAL DYES, Brain, Vol. 81, pp. 248-257, 1958.
9. Eugene J. Agnello & Gerald D. Laubauch. 16-METHYLENE-PREGNENE DERIVATIVES, (to Chas. Pfizer & Co., Inc.), U.S. 2,865,808, December 23, 1958.
10. Amelio E. Montagna, Donald G. Kubler, & Julius J. Brezinski, POLYMERIC COMPOSITIONS & DERIVATIVES FORMED BY THE REACTION OF 2-ALKOXYTETRAHYDROPYRANS & KETENE. (To Union Carbide Corporation). U.S. 2,876,230, March 3, 1959.
11. Frederick Comte, AMINOFORMAMIDO URACILS, (to Monsanto Chemical Company). U.S. 2,852,516, September 16, 1958.
12. A. A. Benson, DYNAMIC INTERRELATIONS OF ENZYME SYSTEMS IN HEXOSE SYNTHESIS IN VIVO. Proceedings of the International Symposium on Enzyme Chemistry in Tokyo and Kyoto, 1957. Vol. 2, pp. 1890191, Publ. 1958 in English.
13. A. N. Stroh, DISLOCATIONS AND CRACKS IN ANISOTROPIC ELASTICITY, Philosophical Magazine, Series 8, Vol. 3, pp. 625-46, 1958.
14. K. H. Pausacker, REACTIONS OF AROYL PEROXIDES. V. BENZOYL PEROXIDE WITH FERROCENE. Australian Journal of Chemistry, Vol. 11, pp. 509-12, 1958.

15. Burton L. Baker, RESTORATION OF INVOLUTED ZYMOGENIC CELLS IN HYPOPHYSECTOMIZED RATS BY REPLACEMENT THERAPY, Anatomical Record, Vol. 131, pp. 389-403, 1958.
16. Philip Person & Albert Fine, OXIDATIVE METABOLISM OF CARTILAGE, Nature, Vol. 183, pp. 610-611, 1959.
17. G. B. Spurr, George Barlow & Hayden Lambert, INFLUENCE OF PROLONGED HYPOTHERMIA & HYPERTHERMIA ON CARDIAC RESPONSE TO INJECTED POTASSIUM, American Journal of Physiology, Vol. 196, pp. 696-702, 1959.
18. James A. Brown & Whitney H. Mears, PHYSICAL PROPERTIES OF PERFLUOROBUTANE, Journal of Physical Chemistry, Vol. 62, pp. 960 - 62, 1958.
19. Patent to Continental Oil Company, LUBRICATING-OIL DETERGENTS, Brit. 806,595, December 31, 1958.
20. Patent to Hercules Powder Company, ORIENTATED FILMS OF SUBSTITUTED OXACYCLOBUTANE POLYMERS, Brit. 802,683, October 8, 1958.
21. David O. DeFree & Wm. R. Eller, MALONIC ACID & ITS ESTERS AND SALTS, (to Ethyl Corporation). U.S. 2,852,559, September 16, 1958.
22. James Cason & Elmer J. Reist, STRUCTURE OF SUCCINYL DICHLORIDE & OF CERTAIN OTHER POTENTIALLY CYCLIC DIFUNCTIONAL COMPOUNDS, Journal of Organic Chemistry, Vol. 23, pp. 1492-6, 1958.
23. Maurice Morton and Irja Piirema, THE BRANCHING REACTION. II. STYRENE AND METHYL METHACRYLATE, Journal of the American Chemical Society, Vol. 80, pp. 5596-602, 1958.
24. D. N. Langenberg & T. W. Moore, CYCLOTRON RESONANCE IN ALUMINUM, Physical Review Letters, Vol. 3, pp. 137-8, 1959.
25. I. Alter, E. Foa, Z. Hadari, G. Peri & J. Trocker, SELECTIVE LEACHING OF URANIUM FROM PHOSPHATE ROCK BY DILUTE MINERAL ACIDS, Proceedings of the U.N. International Conference on the Peaceful Uses of Atomic Energy, 2nd, Geneva, 1958, Vol. 3, pp. 253-9.

ENGINEERING

1. \*J. P. Gordon & L. D. White, NOISE IN MASTER AMPLIFIERS - THEORY & EXPERIMENT, Institute of Radio Engineers, Proceedings, Vol. 46, No. 9, p. 1588-94, September 1958.
2. \*H. C. Pratt, PURGED ENCLOSURES FOR HAZARDOUS ATMOSPHERES, Electrical Manufacturing, Vol. 62, No. 4, p. 112-3, 286; October 1958.
3. T. Kumai, SOME MEASUREMENTS OF ACCELERATION OF HULL VIBRATION AND HUMAN SENSITIVITY TO VIBRATION, European Shipbuilding, Journal of the Ship Technical Society, Vol. 6, No. 3, p. 76-9, 1957.
4. \*E. J. Flynn, C. E. Kilbourne & C. D. Richardson, ADVANCED CONCEPT FOR TURBINE-GENERATOR STATOR-WINDING INSULATION, American Institute of Electrical Engineers - Transactions, Vol. 77, pt. 3 (Power Apparatus and Systems), No. 36, p. 358-71, June 1958.
5. \*B. Wolfe, GENERAL THEORY OF CONTROL SHEETS, Nuclear Science & Engineering, Vol. 4, No. 5, p. 635-48, November 1958.
6. \*R. W. Tuthill, CO<sub>2</sub> WELDING OF STEEL, Tool Engineer, Vol. 40, No. 1, p. 82-86.
7. R. D. Evison, SOME NOTES ON PILE FOUNDATIONS, New Zealand Engineering, Vol. 13, No. 1, p. 32-36, January 15, 1958.
8. By assorted authors, ADVANCING FRONTIERS OF PRODUCTION ENGINEERING, (Contains Abstracts of Papers Presented before 1958 Summer School of Institute of Production Engineers, London.) Mass Production, Vol. 34, No. 10, p. 65-8, 125-6, October 1958.
9. G. Nagelschmidt, E. S. Nelson, E. J. King, D. Attygalle & M. Yoganathan, RECOVERY OF QUARTZ & OTHER MINERALS FROM LUNGS OF RATS-STUDY IN EXPERIMENTAL SILICOSIS, Great Britain Safety in Mines Research Establishment--Report No. 146, 28 pages, May 1958.
10. J. T. McAloren & G. F. Reynolds, ABSORPTIOMETRIC DETERMINATION OF TRACES OF CHROMIUM IN NICKEL & VANADIUM, OF VANADIUM IN CHROMIUM, & OF NICKEL IN CHROMIUM & VANADIUM, Metallurgia, Vol. 57, No. 339, p. 52-6, January 1958.
11. P. R. Ord, OBSERVATIONS ON GLASS CUTTING BY DIAMOND, Society of Glass Technology-Journal, Vol. 41, No. 201, p. 245-58, August 1957.
12. R. D. Robertson, DESIGN CHARTS FOR PRESTRESSED CONCRETE CONTINUOUS BEAMS OF UNIFORM SECTION, Institution of Civil Engineers-Proceedings, Vol. 10, Paper Number 6,231, p. 39-58, May 1958.
13. L. Fried, PREDICTION OF TEMPERATURES IN FORCED-CONVECTION COOLED ELECTRONIC EQUIPMENT, Institute of Radio Engineers-Transactions on Component Parts, Vol. CP-5, No. 2, p. 102-7, June 1958.

14. H. Boppel, AUTOMATIC GAGING KEY TO GREATER PRODUCTIVITY, Tool Engineer, Vol. 40, No. 3, p. 75-8, March 1958.
15. R. C. Taylor, FERRITE CORES FOR COMMUNICATION COILS, Western Union Technical Review, Vol. 12, No. 3, p. 113-118, July 1958.
16. W. A. Summers, CENTRAL STATION CONTROL-TODAY & TOMORROW, Combustion, Vol. 30, No. 1, p. 34-42, July 1958.
17. I. F. Barditch, CATHODE FOLLOWER APPLICATIONS, Electronic Equipment Engineering, Vol. 6, No. 8, p. 75-6, August 1958.
18. K. M. Fuechsel, DESIGN FOR FUTURE HIGH-SPEED GLIDING VEHICLES, Astronautics, Vol. 3, No. 8, p. 34-5, 72-3, August 1958.
19. H. Fujita, HIGHER APPROXIMATION OF RELAXATION SPECTRA FROM DYNAMIC MEASUREMENTS, Journal of Applied Physics, Vol. 29, No. 6, p. 943-6, June 1958.
20. A. P. Young, & D. H. Chandler, PRECISION ELECTRONIC SWITCH FOR FIXED COIL RADAR DISPLAYS, Marconi Review, Vol. 20, No. 126, p. 94-103, 1957.
21. R. J. Allen, RADIOTRACER INSTRUMENTATION SOLVES TOUGH PROCESS PROBLEMS, Instrument Society of America-Journal, Vol. 5, No. 7, p. 38-41, July 1958.
22. R. Halmshaw & C. Hunt, ASSESSMENT OF ELECTRONIC IMAGE-INTENSIFIER FOR WELD RADIOGRAPHY, Welding & Metal Fabrication, Vol. 26, No. 7, p. 249-55, July 1958.
23. F. Neumann, DAMAGING EARTHQUAKE & BLAST VIBRATIONS, Trends in Engineering, (University of Washington - Engineering Experiment Station) Vol. 10; No. 1, p. 4-10, 24-8; January 1958.
24. E. P. Fisher, POWER BRUSHES TAKE ON MAINTENANCE JOB, Iron Age, Vol. 181, No. 3, p. 80-1, January 16, 1958.
25. A. Lesnewich, CONTROL OF MELTING RATE & METAL TRANSFER IN GAS-SHIELDED METAL-ARC WELDING, Welding Journal, Vol. 37, No. 8, p. 343-53, August 1958; p. 418s-425s, September 1958.

GEOLOGY

1. D. B. Hawkins, F. C. Canney & Frederick N. Ward, PLASTIC STANDARDS FOR GEOCHEMICAL PROSPECTING, Economic Geology, Vol. 54, No. 4, p. 738-44 June - July 1959.
2. C. O. Hines & L. R. O. Storey, TIME CONSTANTS IN THE GEOMAGNETIC STORM EFFECT, Journal of Geophysical Research, Vol. 63, No. 4, p. 671-82, December, 1958.
3. Alan E. Beck & Julia M. Beck, ON THE MEASUREMENT OF THE THERMAL CONDUCTIVITIES OF ROCKS BY OBSERVATIONS ON A DIVIDED BAR APPARATUS, American Geophysical Union, Transactions, Vol. 39, No. 6, p. 1111-23, December 1958.
4. Esper S. Larsen Jr., David Gottfried & Marjorie Molloy, DISTRIBUTION OF URANIUM IN THE VOLCANIC ROCKS OF THE SAN JUAN MOUNTAINS, SOUTH-WESTERN COLORADO, In: International Conference on the Peaceful Uses of Atomic Energy, 2nd, Geneva, 1958 - Proceedings, Vol. 2, Survey of Raw Material Resources, p. 509-14, 1958.
5. E. N. Parker, INADEQUACY OF RING-CURRENT THEORY FOR THE MAIN PHASE OF A GEOMAGNETIC STORM, Journal of Geophysical Research, Vol. 63, No. 4, p. 683-89, December 1958.
6. Robert E. Stevenson, GEOLOGY IN AN ACADEMIC YEAR INSTITUTE, GeoTimes, Vol. 4, No. 1, p. 10, July - August 1959.
7. L. T. Aldrich, G. W. Wetherill, G. L. Davis & G. R. Tilton, RADIOACTIVE AGES OF MICAS FROM GRANITIC ROCKS BY Rb-Sr AND K-A METHODS, American Geophysical Union, Transactions, Vol. 39, No. 6, p. 1124-34, December 1958.
8. Howard N. Sloane, USE OF DECAMIRED FILTERS IN CAVE PHOTOGRAPHY, National Speleological Society, Bulletin, No. 20, p. 61-3, November 1958.
9. George E. Siple, GUIDEBOOK FOR THE SOUTH CAROLINA COASTAL PLAIN FIELD TRIP OF THE CAROLINA GEOLOGICAL SOCIETY, Nov. 16-17, 1957, South Carolina State, Development Board, Division of Geology, Bulletin, No. 24, Page 27, 1959.
10. Maynard G. Christenson, DEVELOPMENTS IN TEXAS & OKLAHOMA PANHANDLES IN 1958, American Association of Petroleum Geologists, Bulletin, Vol. 43, No. 6, p. 1235-47, June 1959.
11. Julius Budel, THE "PERIGLACIAL" MORPHOLOGIC EFFECTS OF THE PLEISTOCENE CLIMATE OVER THE ENTIRE WORLD, Translated by H. E. Wright & David Alt, International Geology Review, Vol. 1, No. 3, p. 1-16, March 1959.
12. Leon Knopoff, SCATTERING OF COMPRESSION WAVES BY SPHERICAL OBSTACLES, Geophysics, Vol. 24, No. 1, p. 30-39, February 1959.

13. Marcel Roubault & Rene Coppins, MIGRATION OF URANIUM IN CRYSTALLINE ROCKS & THE POSSIBLE RELATION OF THIS PHENOMENON TO THE GENESIS OF CERTAIN DEPOSITS, In: International Conference on the Peaceful Uses of Atomic Energy, 2nd, Geneva, 1958 Proceedings, Vol. 2, Survey of Raw Material Resources, p. 335-7, 1958.
14. D.E.T. Bidgood & W. B. Harland, ROCK COMPASS: A NEW AID FOR COLLECTING ORIENTED SPECIMENS, Geological Society of America, Bulletin, Vol. 70, No. 5, p. 641-4, May 1959.
15. Bartholomew Nagy & John P. Wourms, Jr., EXPERIMENTAL STUDY OF CHROMATOGRAPHIC-TYPE ACCUMULATION OF ORGANIC COMPOUNDS IN SEDIMENTS: AN INTRODUCTORY STATEMENT, Geological Society of America, Bulletin, Vol. 70, p. 5, p. 655-9, May 1959.
16. W. D. Hicks, EUDIALTYE & EUCOLITE IN CANADA, Canadian Mineralogist, Vol. 6, Part 2, p. 297-8, 1958.
17. D. L. Inman & T. K. Chamberlain, TRACING BEACH SAND MOVEMENT WITH IRRADIATED QUARTZ, Journal of Geophysical Research, Vol. 64, No. 1, p. 41-7, January 1959.
18. N. C. Gerson, THE INTERNATIONAL (GEOPHYSICAL) YEARS - AN APPRAISAL, Franklin Institute, Journal, Vol. 257, No. 3, p. 172-90, March 1959.
20. Jack E. Oliver, LONG EARTHQUAKE WAVES, Scientific American, Vol. 200, No. 3, p. 131-43, March 1959.
21. R. M. Moxham, GEOLOGIC EVALUATION OF AIRBOURNE RADIOACTIVITY SURVEY DATA, In: International Conference on the Peaceful Uses of Atomic Energy, 2nd, Geneva, 1958 Proceedings, Vol. 2, Survey of Raw Material Resources p. 815-19, 1958.
22. A. S. Furncron, THE DISTRIBUTION AND CHARACTER OF STONE FOR AGGREGATE IN GEORGIA, Georgia Mineral Newsletter, Vol. 12, No. 1, p. 1-8, Spring-Summer 1959.
23. W. R. Wagner, CATALOGUE OF DEEP WELL SAMPLES AND GEOPHYSICAL LOGS TO JANUARY 1, 1959, Pennsylvania Geological Survey, Information Circular 16, 34p, 1959.
24. Gilbert Corwin and Helen L. Foster, THE 1957 EXPLOSIVE Eruption ON IWO JIMA VOLCANO ISLANDS, American Journal of Science, Vol. 257, No. 3, p. 161-71, March 1957.
25. L. R. Wilson, A METHOD OF DETERMINING A USEFUL MICROFOSSIL ASSEMBLAGE FOR CORRELATION, Oklahoma Geology Notes, Vol. 19, No. 4, p. 91-3, April, 1959.

MATHEMATICS 1

1. \*G. A. Croes, A METHOD FOR SOLVING TRAVELING-SALESMAN PROBLEMS, Operations Research, Vol. 6, p. 791-812, 1958.
2. \*A. R. Rzhanitsyn, THE SHAPE AT COLLAPSE OF ELASTIC-PLASTIC PLATES\* SIMPLY SUPPORTED ALONG THE EDGES, Technical Report No. 19, Office of Naval Research, Contract Nonr. 562(A2), Nr-064-406, Brown University, Division of Applied Mathematics, Providence, R. I., 48 pages, 1957.
3. Takesi Isiwata, ON THE RING OF ALL BOUNDED CONTINUOUS FUNCTIONS, Science Report of the Tokyo Kyoika Daigaku, Sect A, Vol. 5, p. 293-4, 1957.
4. \*Walter L. Smith, RENEWAL THEORY & ITS RAMIFICATIONS, Journal of the Royal Statistical Society, Series B., Vol. 20, p. 243-302, 1958.
5. \*C. H. Dix, THE NUMERICAL COMPUTATION OF CAGNIARD'S INTEGRALS, Geophysics, Vol. 23, p. 198-222, 1958.
6. \*G. D. Findlay & J. Lambek, A GENERALIZED RING OF QUOTIENTS, I, II, Canadian Mathematical Bulletin, Vol. 1, p. 77-85, 155, 167; 1958.
7. Geroge W. Mackey, QUANTUM MECHANICS & HILBERT SPACE, American Mathematical Monthly, Vol. 64, No. 8, Part II, p. 45-57, 1957.
8. R. D. Tyler, ONE-DIMENSIONAL TREATMENT OF NON-UNIFORM FLOW, Aeronautical Research Council, Report & Memo, No. 2991, 20 pp., 1957.
9. Robert Finn & David Gilbarg, UNIQUENESS & THE FORCE FORMULAS FOR PLANE SUBSONIC FLOWS, Transactions of the American Mathematical Society, Vol. 88, p. 375-379, 1958.
10. R. N. Sen, ON A TYPE OF RIEMANNIAN SPACE CONFORMAL TO A FLAT SPACE, Journal of the Indian Mathematical Society, (NS) Vol. 21, p. 105-114, 1957.
11. Eliahu I. Jury & Frances Mullin, A NOTE ON THE OPERATIONAL SOLUTION OF LINEAR DIFFERENCE EQUATIONS, Journal of the Franklin Institute, Vol. 266, p. 189-205, 1958.
12. P. G. Rooney, LAPLACE TRANSFORMS & GENERALIZED LAGUERRE POLYNOMIALS, Canadian Journal of Mathematics, Vol. 10, p. 177-182, 1958.
13. R. P. Srivastav, ON THE DERIVATIVES OF INTEGRAL FUNCTIONS, Mathematics Student, Vol. 25, p. 11-15, 1957.
14. Edward E. Grace, TOTALLY NONCONNECTED IN KLEINEN CONTINUA, Proceedings of the American Mathematical Society, Vol. 9, p. 818-821, 1958.
15. A. S. Meligy, ON WHITTAKER FUNCTIONS, Journal of the London Mathematical Society, Vol. 33, p. 456-7, 1958.

16. W. A. Thompson, Jr., A NOTE ON P.B.I.B. DESIGN MATRICES, Annals of Mathematical Statistics, Vol. 29, p. 919-922, 1958.
17. William T. Reid, A. PRUFER TRANSFORMATION FOR DIFFERENTIAL SYSTEMS, Pacific Journal of Mathematics, Vol. 8, p. 127-32, 1958.
18. John Lamperti, STATIONARY MEASURES FOR CERTAIN STOCHASTIC PROCESSES, Pacific Journal of Mathematics, Vol 8, p. 127-32, 1958.
19. I. S. Gal, ON THE THEORY OF  $(m, m)$ - COMPACT TOPOLOGICAL SPACES, Pacific Journal of Mathematics, Vol. 8, p. 721-734, 1958.
20. L. O'Riifeartaigh & J. L. Synge, A PROPERTY OF EMPTY SPACE-TIME, Proceedings of the Royal Society of London, Series A, Vol. 246, p. 299-300, 1958.
21. Richard Block, ON TORSION-FREE ABELIAN GROUPS & LIE ALGEBRAS, Proceedings of the American Mathematical Society, Vol. 9, p. 613-620, 1958.
22. J. L. Synge, WHITTAKER'S CONTRIBUTIONS TO THE THEORY OF RELATIVITY, Proceedings of Edinburgh Mathematical Society, Vol. 11, p. 39-55, 1958.
23. Gerald Berman & Robert J. Silverman, NEAR-RINGS, American Mathematical Monthly, Vol. 66, p. 23-34, 1959.
24. Raphael M. Robinson, THE CONVERSE OF FERMAT'S THEOREM, American Mathematical Monthly, Vol. 64, p. 703-10, 1957.
25. Edmund F. Tyler, THE OSCILLATING INBOARD FLAP AT SUPERSONIC SPEEDS, Journal of Aero/Space Sciences, Vol. 26, p. 56-8, 1959.



APPENDIX XVII (con't.)

MEDICINE

1. M. K. Young, Jr., W. L. Brown & L. G. Serralle, EFFECT OF SEROTONIN CREATININE SULFATE UPON GLUCOSE UPTAKE BY RAT DIAPHRAGM, Surgical Research Unit, Brooke Army Medical Center, Fort Sam Houston, Texas, Medew-R-3-8-58, p. 1-4, October 1958.
2. R. Elliott, A METHOD FOR THE INVESTIGATION OF SUSCEPTIBILITY TO INSECTICIDES IN ANOPHELES LARVAE, Transactions of the Royal Society of Tropical Medicine & Hygiene, London, Vol. 52, No. 6, p. 527-34, November 1958.
3. \*H. M. Blackfield, F. A. Torrey, W. J. Morris & B. V. Low-Beer, THE CONSERVATIVE TREATMENT OF HEMANGIOMAS IN INFANTS & CHILDREN, Journal of the International College of Surgeons, Vol. 30, No. 2, p. 255-61, August 1958.
4. \*A. Melamed, SOME ANATOMICAL & ROENTGENOLOGIC CONSIDERATIONS WITH RESPECT TO THE LUMBOSACRAL SPINE: WITH SPECIAL REFERENCE TO RETRODISPLACEMENT, Radiology, Vol. 71, No. 4, p. 548-52, October, 1958.
5. \*W. D. Altus, Q-L VARIABILITY, MMPI RESPONSES, & COLLEGE MALES, Journal of Consulting Psychology, Vol. 22, No. 5, p. 367-71, October 1958.
6. \*H. O. Bames, BAGGY EYELIDS, Plastic & Reconstructive Surgery, Vol. 22, No. 3, p. 264-6, September, 1958.
7. E. M. McConnell, A CASE OF MALIGNANT SACROCCOCCYGEAL TERATOMA, Journal of Pediatrics (St. Louis), Vol. 53, No. 4, p. 489-93, October 1958.
8. H. Riley & A. Spinks, BIOLOGICAL ASSESSMENT OF TRANQUILLISERS II, Journal Pharmacy & Pharmacology, Vol. 10, No. 12, p. 721-40, December 1958.
9. R. O. Hurst, A RAPID METHOD FOR THE LARGE SCALE PREPARATION OF CALF THYMUS DEOXYRIBONUCLEATE, Canadian Journal of Biochemistry & Physiology, Vol. 36.
10. H. Gergman, A NEW TRACTION CATHETER FOR POST-PROSTATECTOMY HEMOSTASIS, Journal of Urology (Baltimore), Vol. 80, No. 5, p. 399-400, November 1958.
11. H. J. Whiteley, STUDIES ON HAIR GROWTH IN THE RABBIT, Journal of Anatomy (London), Vol. 92, No. 4, p. 563-7, October 1958.
12. N. E. Williams & H. G. Bridge, NEPHROTIC SYNDROME AFTER THE APPLICATION OF MERCURY OINTMENT, Lancet, London, Vol. 2, No. 7047, p. 602, Sept. 20, 1958.
13. E. D. Freis, ESSENTIAL HYPERTENSION, Heart Bulletin, Vol. 8, No. 3, p. 52-4, May - June 1959.
14. K. C. Gadre, THE MANAGEMENT OF DEAFNESS II, Indian Journal of Medical Sciences, Bombay, Vol. 13, No. 2, p. 145-153, February 1959.
15. S. Katz, NOCARDIOSIS, G.P., Vol. 19, No. 6, p. 129, June 1959.

16. C. H. Fleurent, SOMME'S MANUSCRIPT REPORT ON A CASE OF FUNGOID DISEASE OF THE GUMS & UPPER JAW (MADAME DE HARVEN), WITH COPIES OF CONSULTATIONS, BY ROUX, BOYER & DUPUYTREN, Annals of the Royal College of Surgeons of England, Vol. 24, No. 4, p. 270-4, April 1959.
17. L. Block, PROTOTYPE STUDY: 25BED HOSPITAL, Modern Hospital, Vol. 92, No. 4, p. 104, passim, April 1959.
18. F. R. Riesenman & M. B. Pettit, CLINICAL EFFICACY OF PROCHLORPERAZINE (COMPAZINE) IN MENTAL ILLNESS, American Journal of Psychiatry, Vol. 115, No. 11, p. 1032-3, May 1959.
19. E. J. Nathaniel & C. D. Clemente, GROWTH OF NERVE FIBERS INTO SKIN & MUSCLE GRAFTS IN RAT BRAINS, Experimental Neurology, Vol. 1, No. 1, p. 65-81, April 1959.
20. J. R. Creut, SOME SPECTROPHOTOFUORIMETRIC OBSERVATIONS ON BLOOD & URINE CATECHOLAMINE ASSAYS, Pharmacological Reviews, Vol 11, No. 2, p. 296-9, Part II, June 1959.
21. P. R. Bonell, UNDERSTANDING SURGICAL PATIENTS, American Journal of Nursing, Vol. 59, No. 8, p. 1148-9, August 1959.
22. H. L. English, AN ALARMING SIDE-EFFECT OF TOFRANIL, Lancet, London, Vol. 1, No. 7085, p. 1231, June 13, 1959.
23. H. S. Tesorier, USES & ABUSES OF SCOPOLAMINE: A CLARIFICATION, Anesthesia & Analgesia; Current Researches, Vol. 38, No. 2, p. 103-8, March-April 1959.
24. D. P. Hall, OUR SURGICAL HERITAGE, American Journal of Surgery, Vol. 97, No. 6, p. 807-9, June 1959.
25. D. I. Williams, THE SKIN & THE MENOPAUSE, Practitioner (London), Vol. 182, No. 1091, p. 574-7, May 1959.

PHYSICS 1

1. H. Maeda, HORIZONTAL WIND SYSTEMS IN THE IONOSPHERIC E REGION DEDUCED FROM THE DYNAMO THEORY OF THE GEOMAGNETIC ~~sq~~ VARIATION III. Journal of Geomagnetism and Geoelectricity, Vol. 9, No. 2, p. 86-93, 1957.
2. \*F. Palmer, MEASUREMENT OF A VARIABLE SURFACE TENSION, American Journal of Physics, Vol. 27, No. 6, p. 431, September 1959.
3. \*R. A. Hefferlin & W. E. Kreger, COLLIMATOR PENETRATION AND SCATTERING EFFECTS, Nuclear Instruments & Methods, Vol. 3, No. 3, p. 149-57, September 1958.
4. \*G. S. Durham & B. A. Soderberg, EVALUATION OF LATTICE INTERACTIONS IN OR NEAR A CUBIC CRYSTAL, Journal of Chemical Physics, Vol. 28, No. 6, p. 1201-6, June 1958.
5. \*L. G. Smith, MEASUREMENTS OF LIGHT MASSES WITH THE MASS SYNCHROMETER, Physical Review, Vol. III, No. 6, p. 1606-19, September 15, 1958.
6. \*W. E. Kock, RELATED EXPERIMENTS WITH SOUND WAVES & ELECTROMAGNETIC WAVES, Proceedings of the Institute of Radio Engineers, Vol. 47, No. 7, p. 1192-201, July 1959. For editor reference, see #4, Engineering.
7. W. A. Newcomb, MAGNETIC DIFFERENTIAL EQUATIONS, Physics of Fluids, Vol. 2, No. 4, p. 362-5, July-August 1959.
8. J. S. Franklin, AUTOMATED UNIVERSAL DISTRIBUTION PHOTOMETER, Illuminating Engineering Society, Vol. 53, No. 12, p. 667-78, December 1958.
9. E. P. Wohlfarth, HARD MAGNETIC MATERIALS, Advances in Physics, Vol. 8, p. 87-224, April 1959.
10. R. B. Regier, W. H. Burgus & R. L. Tromp, RATIO OF ASYMMETRIC TO SYMMETRIC FISSION OF U<sup>233</sup>. AS A FUNCTION OF NEUTRON ENERGY, Physical Review, Vol. 113, No. 6, p. 1589-92, March 15, 1959.
11. C. J. Maiden, THE STRAIN-RATE SENSITIVITY OF  $\alpha$ -URANIUM, Journal of the Mechanics and Physics of Solids, Vol. 7, No. 2, p. 106-13, March 1959.
12. J. J. Becker, DOMAIN BOUNDARY CONFIGURATIONS DURING MAGNETIZATION REVERSALS, Journal of Applied Physics, Vol. 30, No. 3, p. 387-90, March 1959.
13. K. H. Spring, BONDING OPTICAL COMPONENTS TO METAL MOUNTS, British Journal of Applied Physics, Vol. 9, No. 6, p. 242-6, June 1958.
14. R. N. Nixon, ANOMALOUS BAND INTENSITY IN FERMI RESONANCE, Journal of Chemical Physics, Vol. 31, No. 1, p. 258-60, July 1959.

15. A. Rose & M. A. Lampert, GAIN-BANDWIDTH PRODUCT FOR PHOTOCONDUCTORS, R.C.A. Review, Vol. 20, No. 1, p. 57-68, March 1959.
16. W. W. Scales, SPECIFIC HEAT OF LI and KI AT LOW TEMPERATURES, Statistical Review, Vol. 112, No. 1, p. 49-54, October 1, 1958.
17. T. Okuda & K. Yamamoto, DISTURBANCE PHENOMENA IN PROBE MEASUREMENT OF IONIZED GASES, Journal of the Physical Society of Japan, Vol. 13, No. 10, p. 1212-23, October 1958.
19. J. Ferguson, SOME EXPERIMENTS RELATED TO THE PROBLEM OF INTERNAL CONVERSION OF EXCITATION ENERGY IN AROMATIC MOLECULES & CRYSTALS, Journal of Molecular Spectroscopy, Vol. 3, No. 2, p. 177-94, April 1959.
20. L. E. Beghian, G. H. R. Kegel, & R. P. Scharenberg, FAST SODIUM IODIDE SPECTROMETER & ITS APPLICATION TO MILLIMICROSECOND TIME MEASUREMENT, Review of Scientific Instruments, Vol. 29, No. 9, p. 753-7, September 1958.
21. F. A. Hamm, MOTION OF POSITIVE IONS IN SILVER BROMIDE, Journal of Applied Physics, Vol. 30, No. 9, p. 1468-9, September 1959.
22. J. M. Barnes, W. G. Mayer & E. A. Hiedemann, MULTIPLE DIFFRACTION OF LIGHT BY ULTRASONIC VIBRATIONS IN TRANSPARENT SOLIDS, Journal of the Optical Society of America, Vol. 48, No. 9, p. 663-5, September 1958.
23. C. Z. Erzen, AN EXPRESSION FOR EVAPORATION, Bulletin of the Technical University of Istanbul, Vol. 11, No. 1, p. 1-13, 1958.
24. C-S. Yih, INHIBITION OF HYDRODYNAMIC INSTABILITY BY AN ELECTRIC CURRENT, Physics of Fluids, Vol. 2, No. 2, p. 125-30, March-April 1959.
25. B. P. Singh, H. S. Hans & P. S. Gill, THE STUDY OF NOISE PULSES & A LIQUID SCINTILLATOR, Indian Journal of Physics, Proceedings of the Indian Association for the Cultivation of Science, Vol. 32, No. 4, p. 183-91, April 1958.

## APPENDIX XVIII

### AN INTUITIVE COMPUTER LEARNS AN ELEMENT OF GRAMMAR\*

D. G. ELISON

Department of Psychology      Indiana University

Many approaches to problems of information storage and retrieval and mechanical translation include computer programs for classifying words as parts of speech. Existing computer programs with this aim have the form of scientific predictions: classifications are predicted deductively from dictionary and context information by means of grammatical rules that represent hypothesis or knowledge concerning regularities in linguistic behavior. The present study investigates an application of a program which allows the computer to make classifications intuitively rather than deductively. The computer becomes a self-organizing system whose output of decisions is intuitive in the sense that it does not depend upon (the programmer's) prior knowledge or guesses concerning the relevant empirical laws. The program is based upon a statistical model, multiple conditional probability, that is also a theoretical model of human intuitive judgment and of learning. The model and the primitive form of computer in which it is programmed for this study were described earlier (1, 2). The computer which utilizes programs of this type is called EMMA, signifying Empirical Multivariable Matrix Analyzer.

The present investigation is a miniature, a methodological experiment to determine the feasibility of applying the EMMA principle to certain problems of linguistic analysis. The problem of classifying words as parts of speech was chosen from among other linguistic problems in part because conventional

---

\* This research was sponsored by Rome Air Development Center, U.S. Air Force, under contract No. AF 30 (602) - 2185 with Indiana University; F.W. Householder Jr., Principal Investigator, and J. Lyons, Coordinator.

techniques do not provide a simple solution to it and in part because its characteristics are well suited to the requirements of a test of EMMA. There is a clear and practical problem of prediction and, since the classification of words as parts of speech can be done with a minimum of disagreement, questions of criterion reliability may be ignored. Kaplan found (3) that for reducing ambiguity in the meaning of words, a context consisting in two preceding and two following words is approximately as effective as the full sentence in which the word occurs. This suggests that it is not entirely unrealistic to attempt the part-of-speech classification from the four-word context to which we are effectively limited in this study by the small data-handling capacity of the primitive form of EMMA that we chose to use. If more context is necessary we should obtain some notion of the amount required.

EMMA in this study consisted in 500 marginal punched cards, five needles, and pencil and paper, manipulated by a very conscientious young woman, Miss Henrietta Chen. The EMMA principle can be programmed for faster operations upon a larger corpus of data in a conventional electronic computer, but in such computers the operations that mediate the predictions are covert, like those we call mental in man. In this exploratory study it seemed desirable to become thoroughly acquainted with the mediating operations by watching them as they occurred. In our primitive EMMA most of the important ones take place slowly and in the open.

The basic principle of EMMA is multiple conditional probability. Events are considered as points in  $n$ -classification space and stored in a memory as computer words in which a digit-position represents a classification (a mutually exclusive and exhaustive set of classes) and a digit represents a class to which an event is assigned. If information concerning one or more

classifications of an event is absent or ambiguous (as is often the case in this study) the event is stored as a computer word in which more than one digit occurs in a given position; such events appear in the memory as regions rather than as points. Given a partial or ambiguous description of an  $n$ -classificational event, i.e., an event specified by one or more classes in each of  $n - m$  classifications, a prediction of the most likely class of the  $n$ th classification for that event is obtained by searching the memory for computer words that are identical in the classifications that are specified, examining the associated distribution of cases in the  $n$ th classification and (ordinarily) predicting the class containing the greatest frequency. If the number of identical part-words found in the memory is insufficient for reliable prediction the number of cases in the distribution may be increased by examining similar part-words, e.g., words that match in  $n - m - 1$  classifications.

#### Method

In the present application of the EMMA principle, the events are words in a text, each specified in the memory by six classifications, namely the part (or parts) of speech indicated by a dictionary for five consecutive words in a text together with a linguist's classification of part of speech for one of the five. The problem assigned to EMMA was to predict the linguist's classification of a word (presumably a correct classification), given the dictionary classifications of that word and four adjoining words.\*

Using similar programs predictions were made for each word in each of the five positions in sets of five words so that it was possible to compare the accuracy of predictions based on each of five different contexts, there

---

\* Appreciation is due to Miss Beverly Hung, the linguist who, by providing the correct classifications, contributed to EMMA's training.

consisting respectively in four following words, one word preceding and three following, two words preceding and two following, three preceding and one following, and four preceding. Predictions were also made by combining data from two or more context configurations, so that the study provides information concerning the effects of both amount and kind of context on predictive accuracy. The major independent variable in this study, however, is the size of the memory, defined as the number of stored events utilized in making a prediction from a single context. Predictions for a set of 116 words are made with memory sizes of 99, 199, 299, 399, and 499 events. The procedure is closely analogous to that of a human learning experiment in which the accuracy of judgment of some characteristic of complex situations is tested after varying amounts of experience in similar situations for which the correct judgment has been indicated.

The material upon which the program operated was an arbitrarily selected text of some five hundred consecutive words in Karl Stumpff's Planet Earth (4, pp. 23-25).

The new conception, that the fixed stars are distant suns, began to be accepted at about the same time as the heliocentric theory of the solar system. At the end of the eighteenth century Frederick William Herschel, who studied the fixed stars with his giant reflecting telescopes, tried to estimate the size and shape of the stellar system. As a basis for his investigations, he assumed that in the part of space occupied by fixed stars they are distributed in fairly uniform density. By counting all the stars that appeared in certain areas within the range of his most powerful telescope, he created a relative scale for the depth of the stellar system -- at least in his "selective field." By means of these calculations he came to the far-reaching conclusion



that the fixed stars occupy a lens-shaped space of great magnitude, and that the borders of this galaxy are formed by the Milky Way, itself made up of a great cluster of extremely remote stars. He assumed, moreover, that our sun is situated near the center of the Milky Way, whereas present-day astronomers have reached the conclusion that the sun lies far away from this center.

Of course Herschel was not able to make any precise statements concerning the true size and extent of the fixed star system. His telescopes were very powerful for the times, but they were still far from reaching the faint stars which are seen with the biggest modern telescopes. Furthermore, he had only a vague idea of the distance of even the nearest and brightest star. If we assume that the luminosity of the stars is approximately that of the sun -- a supposition that is certainly incorrect as far as individual stars are concerned, and can only be applied to an average of certain type of stars -- then we must conclude that the distance of even the nearest star is enormous when measured by the scale of the solar system.

It was not until 1837, fifteen years after Herschel's death, that Friedrich Wilhelm Bessel, an astronomer from Königsberg, succeeded in measuring accurately the perspective displacement, or parallax, of a fixed star (Fig. 7). Bessel used a newly invented instrument of the greatest precision, the heliometer. It is now known that former experiments had failed because they were directed at the brightest stars, which are not necessarily the nearest. Meanwhile, a further discovery was made: the fact that the fixed stars are not fixed but travel through space, though their motion is hardly perceptible because they are so remote. Knowing this, Bessel sought out a star that

was near enough for the parallax to be observable -- and made his search not among the brightest stars but among those which showed "proper motion." Of these he noticed one in particular, a faint star (No. 61 in the constellation Cygnus) whose annual deviation on either side of the central position is up to 0.33 -- in other words, about one 5,400th part of the angle at which we see the diameter of the full moon. From this, the distance of 61 Cygni could be calculated as more than 60 million million miles: Its light, traveling . . .

#### Coding:

Each word and internal punctuation mark in the selected text was assigned one or more numerals in the following code according to its classification as one or more parts of speech in the Thorndike Century Senior Dictionary. This operation provides Dictionary codes.

1. Noun (including pronouns)
2. Verb (transitive, intransitive and auxiliary)
3. Adjective (including definite and indefinite articles)
4. Adverb
5. Preposition
6. Conjunction
7. Miscellaneous (including all other parts of speech and internal punctuation but not the final period of a sentence)

In carrying out the coding operation, specific questions arose and in each case an arbitrary rule was formulated and followed in subsequent cases of the same kind. Rules were restricted to those that can be programmed on a conventional computer. These rules follow:

1. A final period is not treated as a word (though, since sentences are programmed as separate units, the final period is used

elsewhere in the program).

2. If a word containing affixes, e.g., "stars", does not appear in the dictionary, the stem form ("star") is used and the affixes are ignored. (This rule obviously results in the loss of information contained in affixes, but to use this information it would be necessary to resort to algorithms and a deductive program, which is against our principles in this study, or to increase the amount and complexity of the information used by EMMA in her intuitive predictions, which was not expedient. If EMMA is ever used in practice there would certainly be no objection to supplementing her intuition with logic and any relevant and available empirical laws.)
3. All capitalized words not found in the dictionary are coded as nouns.
4. All numerals are coded as noun (code) and adjective (code 3) since their word equivalents, e.g., "two", are so classified in the dictionary.
5. Multiple-word technical terms such as "fixed star" are treated as two single words, e.g., "fixed" and "star".
6. Hyphenated words are treated as single words if found as combinations in the dictionary; otherwise as two words, ignoring the hyphen.

Each word and internal punctuation mark was also assigned a single code representing its classification as a part of speech by a linguist applying her best judgment to any and all available information. This operation provides the linguist's (or correct) code.

### Punching

Five hundred single-hole marginal punched cards that provided 39 holes on one side were used for the memory. One card was punched for each sequence of five consecutive words in a sentence, word 1 to 5 on the first card, 2 to 6 on the second and so on. Note that the number of cards is four less than the number of words in a sentence, with the result that, for example, the first 100 cards contain information concerning 116 words. There are exactly 100 words in each of the five positions. Seven consecutive holes were assigned to each word, each hole representing one code number. Holes corresponding to the dictionary code (or codes) of each word were notched. The linguist's coding and, for convenience, the word itself, were written below the space assigned to each word. Additional information, useful in sorting operations but not necessary for the program was punched elsewhere.

### Prediction Programs and Procedures

The part-of-speech classifications made by the linguist for the first 116 words in the text were predicted under 45 conditions, corresponding to the 25 combinations of five positions or contexts and five memory sizes, together with 20 additional conditions obtained by combining data from more than one position. The prediction program for words in the first position made with a 99-event memory will serve as an example. The program has three branches.

Branch 1: If the word has a single dictionary code, that code is the prediction.

Branch 2: If the word has more than one dictionary code and appears in the first position (is the first word in a consecutive set of five words in a sentence), the memory is searched for all other events which match that event in any combination of the dictionary codes for its five component words. For these events the distribution of linguist's codes (for the words in the first position) is compiled. The prediction is the modal code in this distribution. If there is more than one mode there is, of course, more than one prediction.

#### Example of procedure in Branch 2:

The first word in the text, "The", appears in the first position on a card carrying the following information: (Each card actually represents five events, one for each position.)

Word	The	new	conception	(comma)	that
Position	1	2	3	4	5
Dictionary codes	3,4	3,4	1	7	1,3,4,6
Linguist's code	3	3	1	7	6

For the event that consists in the dictionary codes and the first-position linguist's code there are 16 possible combinations of the dictionary codes of the five component words, 33171, 43171, 34171, 44171, 33173, etc. A five-needle sort for each combination drops all cards representing events with a matching combination of dictionary codes. The predicted classification for the word "the" is the modal code in the distribution of linguist's codes in the first position on the dropped cards. The card for which a prediction is being made is not included in the distribution so that 100 cards provide a memory of 99 events. It is perhaps obvious that once the sort has been made it is economical of time to examine the distribution of linguist's codes in the other positions, i.e., to obtain the second - position prediction for "new", the third-position prediction for "conception" and so on.

Branch 3: If the word does not appear in the first position (which is the case for the last four words in each sentence, a total of 16 in the first 116 words of the text), the dictionary code or codes is the prediction.

Similar predictions were made for the same 116 words in each of the five positions and the process was repeated after adding 100, 200, 300, and 400 events to the original memory of 99.

Predictions were scored as follows: for each word, if there is a single modal prediction that agrees with the linguist's coding of that word, the score is 1.0. If there are two modal predictions and one agrees, the score is .5, if three and one agrees, the score is .33, etc. If no prediction agrees with the linguist's coding of that word, the score is 0. An average prediction score was obtained for each combination of position and memory size.

Additional predictions were made by pooling distributions of linguist's codes for the same words in more than one position. Combined-position predictions were obtained with each of the five memory sizes with the following combinations of positions: 1,5; 1,3,5; 1,2,3,4,5; and 2,3,4. Predictions were the modal codes as before, and were scored in the same way as those for single-position predictions. Branch 3 of the program was not used in combined-position predictions whenever at least one single-position prediction was obtainable for all words by means of Branches 1 and 2. In the case of the combination of positions 1 and 5, for example, Branch 3 is not used since predictions for the first four words in a sentence are obtained from position 1 data and for the last four words from position 5 data; computer predictions (Branch 2) are possible for the intervening words. In the combination of positions 2, 3, and 4, Branch 3 predictions are made only for the first and last words of each sentence.

A third method of prediction was tried in which single-position predictions rather than distributions were combined, but this was found to be impractical.

#### Results and Discussion

The mean scores for the five single-position predictions obtained with each memory size are shown in Table 1 and Figure 1. The negatively accelerated

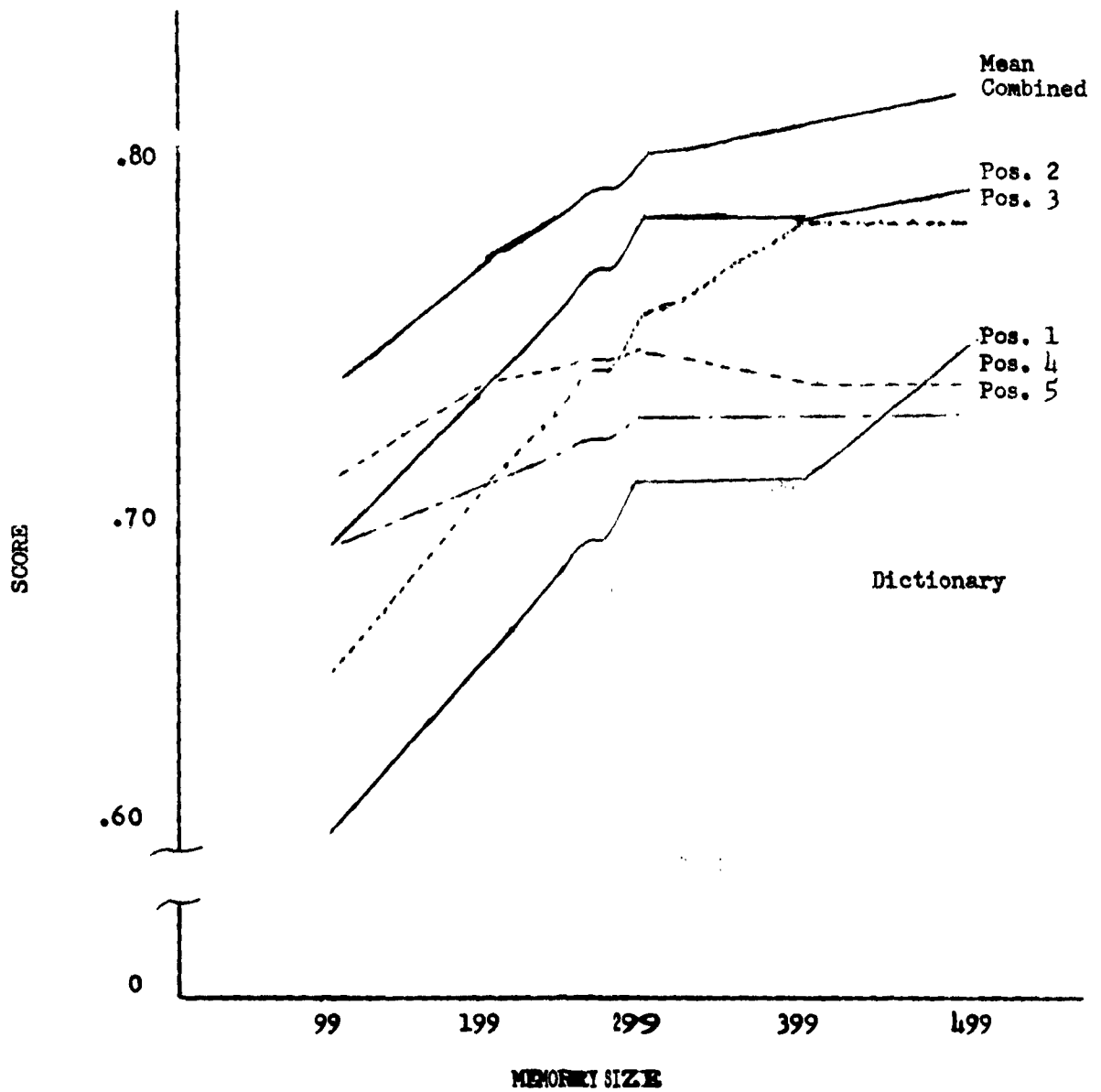


Figure 1

Mean scores for single-position predictions, grand mean scores for combined position predictions and scores for predictions based on dictionary alone.

Table 1

Mean scores for single-position predictions of part-of-speech  
classifications of 116 words

Position	Memory Size				
	99	199	299	399	499
1	.60	.66	.71	.71	.75
2	.69	.74	.79	.79	.80
3	.65	.71	.76	.79	.79
4	.72	.74	.75	.74	.74
5	.69	.71	.73	.73	.73
Grand Means	.67	.71	.75	.75	.76

increase (with sampling deviations) resembles a typical learning curve. As might be expected, accuracy of prediction varies with the context used in making predictions. Although the computer improves upon the predictions obtained from the dictionary (which for the same 116 words scores .67), the maximum accuracy attained, a score of .80 for position 2, is obviously too low for practical use.

The primitive form of EMA used in this study did not permit an increase in the number of context words used in single-position predictive configurations without a corresponding decrease in the fineness of their classification. However, some indication of the effect of increasing the amount of context information upon predictive accuracy was obtained by combining predictions for words in two or more positions. Results obtained for four combined predictions are shown in Table 2 and the grand mean scores are shown graphically in Fig. 1.



Maximum accuracy is increased, to .85 for the combination prediction from

Table 2

Mean scores for combined position predictions of part-of-speech  
classifications of 116 words

Positions	Memory Size				
	99	199	299	399	499
1,5	.69	.74	.76	.78	.81
1,3,5	.70	.76	.81	.83	.85
1,2,3,4,5	.77	.81	.84	.84	.84
2,3,4	.79	.81	.82	.84	.84
Grand means	.74	.78	.81	.82	.83

positions 1, 3, and 5, but visual extrapolation to asymptote gives no reason to suppose that this method of extending the context used in predictions will provide a usable level of accuracy for any memory size. An alternate is to base single-position predictions upon larger context configurations. This will require, of course, a larger memory. To increase the number of context words used in prediction from 4 to 9 while retaining 7-fold classifications increases the number of possible configurations by a factor of 2401, from 16,807 ( $7^5$ ) to 40,353,607 ( $7^9$ ). Because of the restrictions imposed by grammatical rules (assuming that they represent actual linguistic behavior) the number of actual configurations will increase much less rapidly, however. For this and other reasons a commensurate increase in memory size will not be required.

Although the maximum predictive accuracy attained with the limited context information used in this study is not sufficient for practical purposes, the feasibility of the method is demonstrated in principle. The practical question remains, however, whether application to the EMA principle on a larger scale and with modifications of detail will provide predictions of sufficient accuracy. It is to be expected that predictive accuracy will be related not only to the

amount and detail of information used in making predictions, but also to its character. For example, it is obvious that predictive accuracy would be improved by use of the affix information that was ignored in the present study. This information can be incorporated by means of conventional deductive techniques as suggested above, or more simply by applying the EMMA principle as in the present study but utilizing a more detailed dictionary.

#### Summary

A computer program based on conditional probabilities is used to predict the classification of words in text as parts of speech from ambiguous dictionary classifications of those words and a context of four adjoining words. Predictions are based, not on empirical laws (in this case grammatical rules) but on examination of regularities in empirical events (linguistic behavior of a writer) represented in detail in memory. Accuracy of prediction increases as the size of the memory increases and varies also with the character of the context information from which the prediction is made, but the maximum accuracy obtained in this study was not sufficiently high for the requirements of practical use. It is proposed to examine predictions based on larger and more detailed contexts (5).

## References

1. Ellson, D. G. Configurational Analysis, Internal Technical Memorandum No. 61, U. S. Navy Electronics Laboratory, Feb. 1953.
2. Ellson, D. G. A Method for Technological Predictions, in Information Theory in Psychology, H. Quastler, (Ed.) Glencoe, Ill., Free Press, 1955.
3. Kaplan, A. An Experimental Study of Ambiguity and Context. Mechanical Translation, 1955, 2, 39-46.
4. Stumpff, Karl. Planet Earth, Ann Arbor, U. Mich. Press, 1959
5. Entirely independent of the present study is one by K. C. Knowlton of M. I. T., whose results were recently distributed under the title "Sentence Parsing With a Self-Organizing Heuristic Program".

## APPENDIX XIX -- Concerning the Matching Formulae

First, having 2 clusters in common seems more than twice as significant as having 1 in common: likewise for 3 over 2. For instance, if  $n_a = 5$ ,  $n_b = 5$ ,  $n_{ab} = 2$  the words are more nearly synonymous than if  $n_a = 5$ ,  $n_b = 5$ ,  $n_{ab} = 1$ . So the original contribution does not seem to give us a satisfactory ordering for our word pairs. At least as long as we use more or less normal thesauri, taking the  $n_{ab}^{\text{th}}$  root seems to be the simplest modification, giving us the word pair ordering which we think we want. This modification does not change the rankings for fixed  $n_{ab}$  - ex.  $n_a = 5$ ,  $n_b = 10$  still ranks even with  $n_a = 7$ ,  $n_b = 7$  for each corresponding value of  $n_{ab}$ . The effect of the new formula is to raise and flatten the graph for  $n_{ab} = 2$  --more so for  $n_{ab} = 3$ . Thus  $\underline{s}$ , for  $n_{ab} = 3$ ,  $n_a = 5$ ,  $n_b = 5$ , is 4 times as large as for  $n_{ab} = 1$ ; for  $n_a = 7 = n_b$ ,  $n_{ab} = 2$   $\underline{s}$  is 3 times as large as for  $n_{ab} = 1$  and for  $n_{ab} = 3$  is 5 times as large.

Secondly, the subject-predicate division may be inadequate. Predicate adjectives and objects seem to be at least as important as verbs and more important than adverbs. This suggests a division of the sentence into S, V and P-O with adverbs designated  $V_2$ . Unfortunately there is no simple geometric schema for this since a line cannot have more than two sides to its origin. Nevertheless, with the proper equivalences, the same table as before will suffice. Another question: adjective modifying subject is often interchangeable with verb or object - should it always receive a lower rating?

Thirdly, we have a FLEX rating  $\underline{f}$  and semantic rating  $\underline{s}$  giving us a rating for syntactic structure similarity and semantic nearness. We have no rating for relative information content. Supply, component, plant, etc. receive a semantic proximity rating of 1 for a perfect match and a FLEX

rating of 1 if one of them is a subject in both question and text sentence; yet diet matches nutrition with only a .58 semantic rating and even on a  $P_1$  to  $P_2$  flexing get rated only a .63 for a value in the sentence of only .37 compared with the words of low information count which may contribute 1. So, we are experimenting with rating words on information content c. Some success has been achieved with the following content formula

$$c = \frac{2 \sqrt{n_a}}{n_a + 1}$$

which is just  $\frac{\text{geometric mean}}{\text{arithmetic mean}}$  between  $n_a$  and  $n_b = 1$ , the minimum number of semantic cluster memberships. (Thus it is never greater than unity.)

Given a word  $a_i$  in the question sentence and a word  $b_j$  in the text we would compute  $s_{ij}$ ,  $c_i$ ,  $c_j$  &  $f_{ij}$  and then the total rating of the word as to its importance or information value  $v$  relative to the question would be:

$$v_{ij} = c_i c_j s_{ij} f_{ij}$$

Clearly, the new formula does not alter the relative value  $v$  of words, for fixed  $n_a$ ,  $n_b$  - etc. a perfect match for  $n_a = 3$ ,  $n_b = 3$  is still 3 times what it is for  $n_{ab} = 1$ , and  $n_{ab} = 2$  is still  $2\frac{1}{2}$  times as much. However,  $n_{ab} = 1$ ,  $n_a = 2$ ,  $n_b = 3$  gives the same value as  $n_{ab} = 1$ ,  $n_a = 1$ ,  $n_b = 5$ , and  $n_{ab} = 3$ ,  $n_a = 7$ ,  $n_b = 7$  as far as relative information value is concerned.

Fourthly, the denominator is the formula for total match, i.e. the geometric mean of the number of words in the sentence, is so unrelated to our manner of rating that it may be possible for a sentence to be rated above 1. Worse, the explicit answer to the question may easily be a 20 word, unbroken sentence with only a few words corresponding to the question. We are currently experimenting with the following kind of formulae

$$T = \frac{\sum v_{ij}}{\sum v_{ij} + \sum P_{ij}} \quad \text{and} \quad T = \frac{\sum v_{ij}}{\sum v_{ij} + \sum P_{ij} + \sum \max_i (v_{ij})}$$

where  $P_{ij}$  is the computation of the perfect matches in the question with itself and  $\sum \max_i (v_{ij})$  is to be interpreted as the maximum of the perfect flex matches of the question.

We are well aware that accurate evaluations of our corpus sentences with respect to the question is probably not feasible - in fact we almost assume not. We do want to maximize the output ratio  $\frac{\text{information paragraphs}}{\text{junk paragraphs}}$  as well as the ratio  $\frac{\text{information output}}{\text{information in corpus}}$ . It is also important that the 1st ratio be very high at the beginning of the output and that it be low only when the questioner probably has sufficient information. Our experience so far indicates that we often have more than one chance to retrieve a given information paragraph - especially if one of the topic sentences is vaguely worded. Also high information sentences may be more specific in their language than low information sentences. It is also possible to put questions in more than one format.

Even with the above suggestions and even if they work, we run into big junk troubles. When a person asks about navigation on a particular river, how can we avoid giving him information on each of the world's rivers? Or if he wishes to know about a certain kind of lens, how to avoid giving him everything? Pronoun use may really add to the difficulty here.

Especially, though, when the questioner is so specific as to use a name, it seems that we should not even evaluate the sentences of an article which neither use the name nor a match with the name, (if interested in a certain astronaut, his ship would be a match).

A further suggestion is this: (assuming the questions phrased without excess words): find the best semantic match for each word in the question sentence which exists in the article under consideration; then multiply every sentence valuation by a function of those (semantic?) matches - i.e., by a measure of the maximum probable relevance of the article as a whole to the question.

This is another advantage of having the articles or chapters indexed by cluster number rather than having sentence numbers so indexed.

This example illustrates the sensitivity of the sentence ranking to the form of the denominator which is chosen.

R. V. Cook

- Q. Groups of scholars currently working at Rand are carrying out a study of Russian derivational morphology.
- I. Rand Corporation is studying Russian from the viewpoint of word-formation processes.
- N. A group study of Russia has been brought to the attention of scholars.

word	n	c	cluster numbers (Hartrampf + additions)
group	3	.87	146K, 86, 146A
scholar	2	.94	116, 117
work	11	.55	144B, 45, 46B, 31, 13, 126A, 13B, 132D, 144E, 147L, 8
current	5	.78	79, 16, 113, 2, 148E
Rand	1	1	Rand
carry	5	.78	24A, 50, 132D, 24E, 24H
study	4	.80	116, 100B, 114, 119E
morphology	1	1	83
derive	4	.80	49A, 49B, 115, 55B
Russia	1	1	Russia
Corporation	2	.94	146K, 146A
viewpoint	1	1	113
process	1	1	132D
word-form	1	1	83
bring	1	1	24A
attend	10	.58	111, 114, 119G, 38, 119E, 5A, 24A, 46B, 118, 146K



s i f X c <sub>i</sub> c <sub>j</sub>	group S <sub>1</sub>	scholar S <sub>2</sub>	work S <sub>2</sub>	current S <sub>3</sub>	Rand S <sub>3</sub>	carry V <sub>1</sub>	study O <sub>1</sub>	morphology O <sub>2</sub>	derive O <sub>3</sub>	Russian O <sub>3</sub>
Corporation S <sub>1</sub>	.90 X 1 X.81=.73									
Rand S <sub>2</sub>					1 X.32 X 1=.32					
study V <sub>1</sub>		.35 X.47=.16 X.75=.12					1 X.71 X.64=.45			
Russian O <sub>1</sub>										1 X.34 X 1=.34
viewpoint O <sub>2</sub>				.45 X.29=.11 X.78=.04						
process O <sub>3</sub>			.29 X.24=.07 X.55=.04			.45 X.28=.13 X.78=.10				
word- formation O <sub>4</sub>							1 X.17 X 1=.17			
study S <sub>1</sub>	.35 X.63=.22 X.64=.14						1 X.71 X.64=.10			
group S <sub>2</sub>	1 X.63 X.76=.48									
Russian S <sub>3</sub>										1 X.18 X 1=.18
bring V <sub>1</sub>						.45 X X.78=.30				
attention O <sub>1</sub>	.18 X.71=.13 X.45=.06		.10 X.47=.05 X.32=.02			.45 X.71=.10 X.46=.05	.56 X X.46=.25 .35 X.63=.22			
scholar O <sub>2</sub>		1 X.35 X.88=.31					X.75=.16			

-172-

$$\frac{\sum}{\sqrt{m_g \cdot m_c}} \quad I = \frac{2.36}{3.37} = .28$$

$$\frac{\sum}{\sqrt{m_g \cdot m_c}} \quad N = \frac{2.40}{7.75} = .31$$

$$I = \frac{2.36}{4.81} = .49 \quad N = \frac{2.40}{5.47} = .25$$

$$I = \frac{2.36}{7.17} = .33 \quad N = \frac{2.40}{7.71} = .33$$

$$\frac{\sum}{\sum + \sum - \sum_{men}} \quad *$$

$$\frac{\sum}{\sum + \sum}$$

## **DISCLAIMER NOTICE**

**THIS DOCUMENT IS BEST QUALITY  
PRACTICABLE. THE COPY FURNISHED  
TO DTIC CONTAINED A SIGNIFICANT  
NUMBER OF PAGES WHICH DO NOT  
REPRODUCE LEGIBLY.**

/